HAL
open science

Belief revision and incongruity: is it a joke?

Florence Dupin de Saint-Cyr, Henri Prade

To cite this version:

Florence Dupin de Saint-Cyr, Henri Prade. Belief revision and incongruity: is it a joke?. Journal of Applied Non-Classical Logics, 2023, Special issue in honor of Philippe Besnard, 33 (3-4), pp.467-494. 10.1080/11663081.2023.2244379 . hal-04195804

HAL Id: hal-04195804
https://hal.science/hal-04195804v1

Submitted on 4 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Belief Revision and Incongruity: is it a joke?*

Florence Dupin de Saint-Cyr Henri Prade†

IRIT-CNRS, Université de Toulouse,
118, route de Narbonne, 31062 Toulouse Cedex 9, France
August 2023

Abstract

Incongruity often makes people laugh. You have to be smart to say stupid things. It requires to be even smarter for understanding them. This paper is a shameless attempt to formalize this intelligent behavior in the case of an agent listening to a joke. All this is a matter of revision of beliefs, surprise and violation of norms.

Keywords: Humor; belief revision; surprise; incongruity

"Incongruity is never superfluous"
Thea Arbee (2016) *Is the superfluous a (new) modality?*[1]

1 Introduction

Even if much has been written about ingredients that trigger laughter, researchers are still far from having completely understood their interplay in the cognitive process that leads a listener to guffaw at a pun or a joke. They are even farther from a detailed analysis and modeling of the mechanisms that are at work in this process.

However, in recent articles Dupin de Saint-Cyr and Prade (2020, 2022) took a first step in this direction by laying bare that a belief revision mechanism was solicited in the reception of a narrative joke. Namely the punchline, which triggers a revision, is both *surprising* and *explains perfectly* what was reported in the beginning of the joke. A similar idea has been more informally proposed in Ritchie (2002). It is quite clear that this is insufficient for characterizing a narrative joke. One may experience similar surprise and revelation effects when reading a detective story or when following the path of a scientific discovery, while there is usually nothing to laugh about.

This latter observation suggests that there are other ingredients in a joke. Among possible candidates that may play a role, incongruity has been often identified. Incongruity has a rather broad meaning, since it may cover anything contrary to custom or good manners. This ranges from unexpected behavior to absurd or crazy reasoning to violations of taboos. This is a rather large scope of possibilities which have in common the non-respect of some (social) norms. In this paper we look at how to incorporate this ingredient into our model after rediscussing the defeasible setting in which it takes place.

The rest of the article is structured as follows. Section 2 provides an overview of the main ingredients of fun, as discussed by philosophers and psychologists, emphasizing the role of incongruity. Section 4

*A special paper on/in humor/honor for/of Philippe Besnard.

This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Applied Non-Classical Logics on August 2023, available at: https://doi.org/ 10.1080/11663081.2023.2244379.

†Contact: `florence.bannay@irit.fr` and `henri.prade@irit.fr`

[1]Thea Arbee aka Léa Sombé ("Les A sont B" / "The A are B") is the pen name of a group of researchers active in the early 1990's to which Philippe took part. It happened that Léa Sombé embodied herself during brief appearances to celebrate members of the group. Léa Sombé sends her best kisses to Philippe.

restates, details and exemplifies the authors'model based on belief revision and surprise, using nonmonotonic reasoning methods, as recalled in Section 3. Section 5 discusses how to add incongruity by taking into account norms. The conclusion points out future lines of research for a better coverage of the multiple aspects involved in humor.

2 Incongruity among other ingredients of fun

In the philosophical and psychological literature on humor, three ideas have often been put forward to trigger laughter: surprise, incongruity and superiority (see Keith-Spiegel (1972) for an historical account).

- *Surprise*: Surprise has been considered as a necessary condition for laughter by many authors and for a long time. Descartes (1649) in *Les Passions de l'Âme* (Passions of the Soul) (sections CXXIV to CXXVI) analyzes the physiological aspect of laughter and its causes by underlining that it results from a surprise (related to "Admiration") occurring in a context of (moderate) "Joye". In a psychological study, Hollingworth (1911) shows that when novelty or surprise is eliminated then the response to a humorous stimulus is diminished. The importance of surprise in the analysis of humor is acknowledged by Marius Latour (1956) who dedicated the fourth essay of his 1956 book on laughter to "la surprise source du rire" (surprise as a source of laughter).

- *Incongruity*: Incongruity is also advocated as a necessary ingredient of humor, e.g., Ritchie (1999). It is often associated to surprise. Beattie (1776) notes very early that humor can be based on ideas/situations that are divergent from normal customs, or based on inconsistent and inappropriate situations. Kant (2001), as Schopenhauer (1819), speaks about the sudden incongruity of the perception of a real object compared to what had been imagined/expected. In his book *Laughter: An Essay on the Meaning of the Comic*, Bergson (1900), advocates that the source of laughter involve an over-rigid response, that would, to use Bergson's terms, "superimpose the mechanical on the living"; this may be viewed as a particular case of incongruity. Delage (1919) sees incongruity as a disharmony which is an essential condition of the comic: "Il doit exister entre l'effet produit et sa cause une désharmonie quantitative ou qualitative, d'où résulte une impression de surprise, un effet d'imprévu" (There must exist between the effect produced and its cause a quantitative or qualitative disharmony, from which results an impression of surprise, an effect of the unexpected). Other references and more details on the theory of incongruity can be found in the collection of papers edited by Chapman and Foot (1976).

- *Superiority* : According to F. R. Fleet (1890), referring to the *Leviathan* Hobbes (1651), "Laughter is a sudden glory arising from some sudden conception of some eminency in ourselves by comparison with the inferiority of others". Thus what is laughable would be the highlighting of an imperfection (absent in the listener) by the one who wants to make people laugh, arousing a sense of superiority. The authors of this theory see laughter as a victory over other people (or circumstances). One feels superior by laughing because one is less stupid, less ugly, less unlucky, less weak than the target of mockery. One may consider that the notion of superiority underlies Bergson (1900)'s famous book on laughter, notably through the theatrical aspect of his examples (where the audience is laughing at the characters embodied by the actors). In the same vein, Marcel Pagnol (1947), a well-known French play-writer has written a short essay on this topic where he emphasized that "Laughter is a song of triumph (chant de triomphe): it is the expression of a momentary superiority, but suddenly discovered, of the laugher on the mocked. There are two kinds of laughter", one considered by Pagnol as positive ("healthy, invigorating, relaxing") where one "feels superior to the whole world or to oneself", the other negative which is "hard and almost sad", where one laughs because the other is inferior to oneself.

These three elements have also been mentioned by other authors, see Attardo and Raskin (1991); Ritchie (2018), sometimes together with more ingredients such as "double entendre", "lack of compassion", "situation matching", etc. For instance, Maier (1932) summarises his view as: "The thought-configurations which makes for a humorous experience must (1) be unprepared for; (2) appear suddenly and bring with it

a change in the meaning of its elements; (3) be made up of elements which are experienced entirely objectively [...] ; (4) contain as its elements the facts appearing in the story, and these facts must be harmonized, explained and unified; and (5) have the characteristics of the ridiculous in that its harmony and logic apply only to its own elements" [pp.73-74].

In this paper, we would like to continue our exploration about joke analysis by taking into account incongruity as a new part of the mechanism introduced in Dupin de Saint-Cyr and Prade (2020) where surprises were handled in a logical setting involving nonmonotonic reasoning. We leave the superiority ingredient for further research since it involves a richer framework able to represent the beliefs of several agents (the teller, the listener and the other people that are mocked). We note that the notion of "incongruity resolution" plays an important role in psychology literature, see Bariaud (1983); Chapman and Foot (1976); Ritchie (1999): this expression means finding a logic or at least find the initial thing less incongruous than it seemed, i.e., "make it acceptable" (inasmuch the matter is not perceived as serious or dangerous, in particular with a young audience). The French psychologist Françoise Bariaud's essay about the genesis of humor in children (Bariaud (1983)) claims that the notion of cognitive reference frame (which the child is in the process of constructing) is used both to detect the incongruity and to resolve it (calling upon a logical sense). She also notes that it is not necessary to put taboos into play, as we can exemplify by the following riddle: *How to put four elephants in a Citroën 2CV ?* Answer: *two in the front and two in the back.*

It seems to us that there is often a latent mishmash between inconsistency and incongruity. We believe that incongruity is an important ingredient which by its comic nature facilitates laughter, and thus adds to the surprise of the contrast (or even inconsistency) of the punchline with the expectation created by the context. Incongruity[2] usually involves the *violation of norms* which can provoke a comic effect (insofar as the violation is not too shocking for the receiver). Hence incongruity also relates to inconsistency with respect to particular pieces of knowledge considered as "norms" (such are rules of good behavior) which are disregarded in the joke. Incongruity could be viewed as a spice, chosen in a more or less appropriate way, which is added in greater or lesser quantity to a dish according to the guests to whom it is served. Thus, for certain audiences, certain jokes will be too salty, or too bland.

We now recall several AI notions that are at work in our modeling of joke understanding, including our previous proposal, see Dupin de Saint-Cyr and Prade (2020).

3 Background on belief revision and nonmonotonic reasoning

With humor, there is a manipulation of what listeners are meant to believe in order to make them fall into a trap Raccah (2015). In this section, we simply propose to translate this idea in terms of belief revision as also suggested by Ritchie (2002) who introduced a logical language to formalize the interpretation of jokes. Ritchie started from the principle that "the punchline forces to reinterpret the context previously set up". His language includes an accommodation operator (used to integrate non-conflicting knowledge) and a belief change operator to take into account the punchline, this operator is inspired by revision in the sense of Gärdenfors (1988).

Moreover the comprehension of a joke by a listener involves modeling her or his commonsense knowledge with which she or he interprets the story told. This is a matter of nonmonotonic reasoning. This background section first provides refreshers on belief revision, and then on nonmonotonic reasoning, before interfacing the two settings in the framework of possibility theory. A brief review on the modeling of surprises is also added.

3.1 Belief revision refresher

We temporarily abandon the theme of laughter to recall the notion of belief revision which is central in our analysis of the understanding of jokes. Alchourrón, Gärdenfors, and Makinson (1985) (AGM) have introduced the concept of belief revision. Belief revision is the process of determining what remains of old beliefs after the arrival of new information. The beliefs are represented by expressions in a formal language. *Revision* consists in adding information while preserving consistency. This operation is necessary since

[2]Incongruity is lacking congruity which comes directly from late Latin *congruus* "suitable, agreeing," Dictionary (2023)

inconsistency leads to a useless state of belief. The main contribution of the AGM article is the definition of a set of postulates that must be satisfied by any *rational* revision operator. As emphasized by Gärdenfors (1988) - see also Léa Sombé (research group) (1994) - these postulates are based on three principles:

- a principle of consistency (the result must be consistent),
- a principle of minimal change (the initial beliefs must be modified as little as possible),
- a principle of priority to the new information (the new information must hold after the revision).

We consider a propositional language $\mathcal{L}$, where propositions are denoted by Greek symbols in lower case. The symbols $\perp$, $\vee$, $\wedge$, $\neg$, $\rightarrow$, $\equiv$, $\models$ denote respectively the contradiction, the logical connectors "or", "and", "not", material implication, logical equivalence, satisfiability. Let Ω denote the set of interpretations induced by $\mathcal{L}$, we will often use ω for naming a particular interpretation in Ω, each interpretation will be described by the list of literals satisfied by it , e.g. $\omega = a \neg b\ c$ is an interpretation that associates the truth value True to a and c and False to b. $Mod(A) \subseteq \Omega$ is the set of interpretations satisfying the set of propositional formula $A \subseteq \mathcal{L}$ ($Mod(A) = \{\omega \in \Omega | \omega \models \bigwedge_{\varphi \in A} \varphi\}$), the same notation is used to represent the set of models of a formula $Mod(\varphi) = \{\omega \in \Omega | \omega \models \varphi\}$. In what follows, we use the symbol $\circ$ for representing an operator of *belief revision* in the sense of Katsuno and Mendelzon (1991).

We recall below the set of postulates stated by Katsuno and Mendelzon (1991) which are equivalent to those of AGM, but allow more easily to link a revision operator to a distance relation between interpretations. More formally, a KM revision operator associates to a formula[3] κ and to a formula φ (representing the new information), another formula denoted by $\kappa \circ \varphi$. To be considered as "rational" the operator $\circ$ must satisfy the KM postulates:

(KM1): $\kappa \circ \varphi \models \varphi$
(KM2): If $\kappa \wedge \varphi$ satisfiable, then $\kappa \circ \varphi \equiv \kappa \wedge \varphi$
(KM3): If φ is satisfiable, then $\kappa \circ \varphi$ is also satisfiable
(KM4): If $\kappa_1 \equiv \kappa_2$ and $\varphi_1 \equiv \varphi_2$ then $\kappa_1 \circ \varphi_1 \equiv \kappa_2 \circ \varphi_2$
(KM5): $(\kappa \circ \varphi) \wedge \psi \models \kappa \circ (\varphi \wedge \psi)$
(KM6): If $(\kappa \circ \varphi) \wedge \psi$ is satisfiable then $\kappa \circ (\varphi \wedge \psi) \models (\kappa \circ \varphi) \wedge \psi$

The first four postulates are the basic postulates of revision. The last two agree with the principle of minimization of change, according to Katsuno and Mendelzon (1991); similar postulates had been introduced before as a basis for rational choice functions, see Sen (1971). More precisely, **(KM1)** imposes that the new information must hold after the revision. **(KM2)** dictates that when the new information is not inconsistent with the original beliefs then the revision is a simple expansion. **(KM3)** expresses that if the new information is consistent then the revised set of beliefs is consistent. **(KM4)** expresses that a revision operator is syntax independent. According to Katsuno and Mendelzon (1991), "rule **(KM5)** says that our notion of closeness is well-behaved in the sense that if we pick any interpretation ω which is closest to $Mod(\kappa)$ in a certain set, namely $Mod(\varphi)$, and ω also belongs to a smaller set, $Mod(\varphi \wedge \psi)$, then ω must also be closest to $Mod(\kappa)$ within the smaller set $Mod(\varphi \wedge \psi)$." Moreover Katsuno and Mendelzon (1991) explained that "a violation of rule **(KM6)** would imply that an interpretation ω may be closer to the knowledge base than ω' within a certain set, while ω' is closer than ω within some other set"[4].

Katsuno and Mendelzon stated the following representation theorem for expressing a revision in terms of proximity of models:

Theorem 3.1 (Katsuno and Mendelzon (1991))**.** *$\circ$ satisfies the postulates (KM1) - (KM6) if and only if there exists a function faithfully assigning to each epistemic state κ a total pre-order $\preceq_\kappa$ such that:*

$$Mod(\kappa \circ \varphi) = \min(Mod(\varphi), \preceq_\kappa)$$

[3] In the initial approach of AGM, a deductively closed set K of formulas was considered to represent the initial knowledge, KM showed that one could state the postulates on a formula κ whose set of consequences forms K.

[4] Katsuno and Mendelzon (1991) detailed this argument in the following way: "consider a model ω of $\kappa \circ (\varphi \wedge \psi)$, that is, a model of $\varphi \wedge \psi$ that is closest to $Mod(\kappa)$ within the set $Mod(\varphi \wedge \psi)$. Suppose ω is not a model of $(\kappa \circ \varphi) \wedge \psi$. The precondition of **(KM6)** says that there is some interpretation ω' that is a model of $\kappa \circ \varphi$ and also of ψ. That is, ω' is a model of $\varphi \wedge \psi$ that is closest to $Mod(\kappa)$ within the set $Mod(\varphi)$. But then ω' is closer to $Mod(\kappa)$ within the set $Mod(\varphi)$ than ω, while ω is closer to $Mod(\kappa)$ than ω' within the set $Mod(\varphi \wedge \psi)$."

The assignment is faithful when it associates to any formula κ a pre-order $\preceq_\kappa$ which strictly privileges the interpretations satisfying κ to the others[5]. We recall that a *pre-order* is a reflexive and transitive relation; a pre-order $\preceq$ over a set A is *total* when for any pair (a, b) of elements of A either $a \preceq b$ or $b \preceq a$. Note that in the context of belief revision, the piece of information φ is a new piece of knowledge about the world which is considered as static. When φ is representing an evolution of the world and not an evolution of the knowledge about the world, then the operation is called *update* Katsuno and Mendelzon (1992); Winslett (1988). Belief update has its own postulates and representation theorem, but this issue goes beyond the modeling needs of this paper, since in jokes, the aim of the teller is to manipulates the beliefs of the listener.

As noticed above a revision operator $\circ$ is defined by a family of pre-orders $\preceq_K$, one pre-order for each initial formula K, however in the following we only require to dispose of *one* initial formula characterizing the state of knowledge of the listener before hearing the joke. This is why we only need one pre-order associated to K, namely the listener can be represented by the pair $(K, \preceq_K)$. Still in the definitions of next subsection, we shall use the notation $(K, \circ)$ and benefit of concise writing such as $K \circ \varphi$. As explained in the next subsection, when the knowledge base of the listener includes default rules, it gives birth to a pre-order on interpretations. In other words, a set of default rules is an implicit way to encode $(K, \preceq_K)$.

3.2 Nonmonotonic reasoning and possibility theory refresher

As recalled by the above theorem, the semantics of belief revision is a matter of pre-orders which can be also expressed in terms of sphere systems (Grove (1988)) and epistemic entrenchment (Gärdenfors (1988)), two notions which have their exact counterpart in possibility theory (in the form of qualitative possibility distributions and necessity relations, see Dubois and Prade (1991)). A possibility distribution is qualitative in the sense that it only plays the role of an ordinal ranking function on the set of possible interpretations (unlike a probability distribution). Possibility distributions are used here to implement the revision operator $\circ$ because what is required is simply an order of plausibility as in sphere systems and no numerical value is either required or available. However, the approach presented in this paper is valid for any type of belief revision operator satisfying the Alchourrón et al. (1985) or Katsuno and Mendelzon (1991) postulates.

Gärdenfors (1990) has pointed out that belief revision and nonmonotonic reasoning can be seen as the two sides of the same coin. Indeed, a pre-order is naturally induced by default rules which express generic knowledge. The interest of default rules is the capability of reasoning in presence of incomplete information. In such a setting non-monotonic reasoning takes place for handling exceptions. In our encoding of common sense knowledge, we use $\rightsquigarrow$ to represent a default rule: $\alpha \rightsquigarrow \beta$, with $\alpha, \beta \in \mathcal{L}$; which means that when α is true, it is more plausible that β is true than false.

A procedure, called System Z, that stratifies a set of default rules (according to their ability to tolerate each other) was first proposed by Pearl (1990) (in agreement with an infinitesimal probability setting). This stratification provides ways to obtain pre-orders on interpretations as described in Benferhat, Cayrol, Dubois, Lang, and Prade (1993), among them is the so-called *best-out ordering* that prefers interpretations where the priority level of the highest priority formula violated is as small as possible. Another well-known ordering on interpretations based on stratified sets of formulas is the so-called *lexicographic ordering* that labels each interpretation with a tuple listing the number of violated formulas per strata and compare the obtained tuples by lexicographic order, see Benferhat et al. (1993).

The best-out ordering on interpretations corresponds to the one obtained by Benferhat, Dubois, and Prade (1992) in a possibility theory setting. Moreover in Benferhat et al. (1992) a stratification of the set of default rules is also obtained, which is the same as the one obtained by System Z (Pearl (1990)); see Benferhat, Dubois, and Prade (1997). More precisely, a default rule is translated in the framework of possibility theory (Dubois and Prade (1988)) by the constraint $N(\beta|\alpha) > 0$. Because of the duality between a necessity measure N and its associated possibility measure Π, i.e., $N(\alpha) = 1 - \Pi(\neg\alpha)$, the

[5]More formally, the assignment is faithful if for any formula κ, we have:

1) If $\omega \models \kappa$ and $\omega' \models \kappa$ then $\omega =_\kappa \omega'$.

2) If $\omega \models \kappa$ and $\omega' \not\models \kappa$ then $\omega \prec_\kappa \omega'$.

3) If $\kappa \equiv \kappa'$ then $\preceq_\kappa = \preceq_{\kappa'}$.

constraint $N(\beta|\alpha) > 0$ is written in an equivalent way as $\Pi(\alpha \wedge \beta) > \Pi(\alpha \wedge \neg\beta)$ which expresses that, in the context α, having β true is the normal situation (since strictly more possible than β false). This writing assumes the existence of a plausibility relation on the interpretations, which is represented by a possibility distribution π, i.e., a function from the set of interpretations Ω of $\mathcal{L}$ to a linearly ordered scale bounded by 0 and 1, where 0 represents the impossible and 1 the completely possible. Given a set Δ of n default rules $\alpha_i \rightsquigarrow \beta_i$, encoded as a collection of constraints $\Pi(\alpha_i \wedge \beta_i) > \Pi(\alpha_i \wedge \neg\beta_i)$, a unique possibility distribution π_Δ over interpretations is obtained by selecting the least restrictive[6] possibility distribution satisfying all the constraints (where $\Pi(\varphi) = \max_{\omega \models \varphi} \pi_\Delta(\omega)$). This is how the set of defaults Δ is associated with a pair $(K, \preceq_K)$ such that $K = \bigcup_i \neg\alpha_i \vee \beta_i$ and $\preceq_K = \preceq_\Delta$ where $\preceq_\Delta = \{(\omega, \omega') | \pi_\Delta(\omega) \geq \pi_\Delta(\omega')\}$ which is faithful with respect to K by construction (as required in the Theorem 3.1). Note that the formulas in K can be ranked by using their necessity levels $N(\neg\alpha_i \vee \beta_i)$ induced from π_Δ by $N(\varphi) = \min_{\omega \not\models \varphi}(1 - \pi_\Delta(\omega))$.

Now given a set of default rules Δ representing the initial knowledge, the arrival of a new piece of information φ is integrated by a revision process (hidden behind the notation "$K \circ \varphi$") where $(K, \preceq_K)$ are defined as above, based on Δ. Note that this revision is equivalent to the conditioning of π_Δ by α (i.e., $\pi_\Delta(\omega|\varphi) = \pi_\Delta(\omega)$ if $\omega \models \varphi$, and 0 otherwise), yielding a revision operator (Benferhat, Dubois, Prade, and Williams (2002)) satisfying the revision postulates. This operator is similar to that of Lehmann (1995).

As noticed in subsection 3.1, the listener who could be equated with a pair $(K, \circ)$ can be more simply identified with a pair $(K, \preceq_K)$, K being given. Such an ordering $\preceq_K$ is naturally available when the knowledge is expressed by a set of formulas including default rules (and maybe strict rules). Understanding a joke being a matter of default reasoning, as explained in the introduction of Section 3, it leads to use default rules to represent the knowledge of the listener. This way of encoding the agent's *non-monotonic* knowledge does not affect the generality of our approach. It is simply a compact and explicit way of specifying the plausibility relations between interpretations (which according to Gärdenfors (1990) underlie any revision operator).

3.3 From default reasoning to belief revision

From now on, the listener's knowledge can also be expressed by a knowledge base Σ made of two parts: $\Sigma = (P, \Delta)$ which is the pair of a *consistent* set P of propositional formulas and a *consistent*[7] set Δ of default rules of the form $\alpha_i \rightsquigarrow \beta_i$, with $\alpha_i, \beta_i \in \mathcal{L}$. Indeed, as it is the case in classical logic, an inconsistent knowledge base leads to an impossibility of rational reasoning, thus here the impossibility of understanding the joke. In this paper, Σ plays the role of both K and $\circ$ since Δ allows to induce a pre-order on the interpretations. As seen in subsection 3.2, Δ and $(K, \circ)$ are the two sides of the same coin since $Mod(K \circ \varphi) = \min(Mod(\varphi), \preceq_\Delta)$. With the introduction of the set of strict rules P it is necessary to use a revision under the integrity constraint P in the sense of Katsuno and Mendelzon (1991) defined by $K \circ^P \varphi = K \circ (P \cup \{\varphi\})$ which yields $Mod(K \circ^P \varphi) = \min(Mod(P \cup \{\varphi\}), \preceq_\Delta)$ and which amounts to be equivalent with $\min(Mod(\varphi), \preceq_\Sigma)$ where $\preceq_\Sigma$ is the pre-order obtained from $\preceq_\Delta$ restricted to the models of P. More precisely, from the set Σ, we can compute a possibility distribution π_Σ over the interpretations that satisfy the constraints induced by Δ and assigns 0 to each interpretation that does not satisfy P (Benferhat et al. (1997)).

The aim of the following example is to show how the pair $(K, \preceq_K)$ can be obtained from a knowledge base $\Sigma = (P, \Delta)$. This is done in terms of an apparently abstract example, whose data corresponds exactly to the situation of the joke Example 4.11.

Example 3.2. Let us consider the following knowledge base Σ representing the knowledge of a listener, we are going to show how this listener can rank order the interpretations based on Σ.

$$\Sigma = \left| \begin{array}{ll} P & \neg k \rightarrow \neg c \\ \hline \Delta & \begin{array}{l} t \rightsquigarrow c \\ c \rightsquigarrow \neg r \\ \neg c \rightsquigarrow r \end{array} \end{array} \right.$$

[6] In terms of letting the possibility degrees as large as possible with respect to the constraints.

[7] The consistency of a default base is equivalent to the consistency of the set of constraints of the form $\Pi(\alpha_i \wedge \beta_i) > \Pi(\alpha_i \wedge \neg\beta_i)$ (Benferhat et al. (1997)).

In this example the language is based on four propositional variables: k, t, c, r yielding 16 interpretations that we denote by $\omega_1, \ldots, \omega_{16}$ as follows:

$$\begin{array}{lcl@{\qquad}lcl}
\omega_1 & = & k\ t\ \ c\ \ r & \omega_9 & = & \neg k\ t\ \ c\ \ r \\
\omega_2 & = & k\ t\ \ c\ \neg r & \omega_{10} & = & \neg k\ t\ \ c\ \neg r \\
\omega_3 & = & k\ t\ \neg c\ r & \omega_{11} & = & \neg k\ t\ \neg c\ r \\
\omega_4 & = & k\ t\ \neg c \neg r & \omega_{12} & = & \neg k\ t\ \neg c \neg r \\
\omega_5 & = & k \neg t\ c\ \ r & \omega_{13} & = & \neg k \neg t\ c\ \ r \\
\omega_6 & = & k \neg t\ c\ \neg r & \omega_{14} & = & \neg k \neg t\ c\ \neg r \\
\omega_7 & = & k \neg t \neg c\ r & \omega_{15} & = & \neg k \neg t \neg c\ r \\
\omega_8 & = & k \neg t \neg c \neg r & \omega_{16} & = & \neg k \neg t \neg c \neg r
\end{array}$$

From Σ, applying the mechanism recalled in Section 3.2,

1. Δ is transformed into the three constraints:
   $\Pi(t \wedge c) > \Pi(t \wedge \neg c)$
   $\Pi(c \wedge \neg r) > \Pi(c \wedge r\}$
   $\Pi(\neg c \wedge r) > \Pi(\neg c \wedge \neg r)$.
   They are equivalent to:
   $\max(\pi_\Delta(\omega_1), \pi_\Delta(\omega_2), \pi_\Delta(\omega_9), \pi_\Delta(\omega_{10})) > \max(\pi_\Delta(\omega_3), \pi_\Delta(\omega_4), \pi_\Delta(\omega_{11}), \pi_\Delta(\omega_{12}))$
   $\max(\pi_\Delta(\omega_2), \pi_\Delta(\omega_6), \pi_\Delta(\omega_{10}), \pi_\Delta(\omega_{14})) > \max(\pi_\Delta(\omega_1), \pi_\Delta(\omega_5), \pi_\Delta(\omega_9), \pi_\Delta(\omega_{13}))$
   $\max(\pi_\Delta(\omega_3), \pi_\Delta(\omega_7), \pi_\Delta(\omega_{11}), \pi_\Delta(\omega_{15})) > \max(\pi_\Delta(\omega_4), \pi_\Delta(\omega_8), \pi_\Delta(\omega_{12}), \pi_\Delta(\omega_{16}))$

2. The least restrictive possibility distribution satisfying the constraints is such that $\pi_\Delta(\omega_2) = \pi_\Delta(\omega_6) = \pi_\Delta(\omega_7) = \pi_\Delta(\omega_{10}) = \pi_\Delta(\omega_{14}) = \pi_\Delta(\omega_{15}) > \pi_\Delta(\omega_1) = \pi_\Delta(\omega_3) = \pi_\Delta(\omega_4) = \pi_\Delta(\omega_5) = \pi_\Delta(\omega_8) = \pi_\Delta(\omega_9) = \pi_\Delta(\omega_{11}) = \pi_\Delta(\omega_{12}) = \pi_\Delta(\omega_{13}) = \pi_\Delta(\omega_{16})$

3. We compute π_Σ from π_Δ taking into account the constraint $\neg k \rightarrow \neg c$ of P, by making its counter-example at the lowest level: $\pi_\Sigma(\omega_2) = \pi_\Sigma(\omega_6) = \pi_\Sigma(\omega_7) = \pi_\Sigma(\omega_{15}) > \pi_\Sigma(\omega_1) = \pi_\Sigma(\omega_3) = \pi_\Sigma(\omega_4) = \pi_\Sigma(\omega_5) = \pi_\Sigma(\omega_8) = \pi_\Sigma(\omega_{11}) = \pi_\Sigma(\omega_{12}) = \pi_\Sigma(\omega_{16}) > \pi_\Sigma(\omega_9) = \pi_\Sigma(\omega_{10}) = \pi_\Sigma(\omega_{13}) = \pi_\Sigma(\omega_{14}) = 0$

4. This corresponds to the pair $(K, \preceq_K)$ where $K = \{\neg t \vee c, \neg c \vee \neg r, c \vee r\}$ and $\preceq_K = \preceq_\Sigma^{bo}$ is such that[8] $\{\omega_2, \omega_6, \omega_7, \omega_{15}\} \preceq_\Sigma^{bo} \{\omega_1, \omega_3, \omega_4, \omega_5, \omega_8, \omega_{11}, \omega_{12}, \omega_{16}\}$. The set $\{\omega_9, \omega_{10}, \omega_{13}, \omega_{14}\}$ being excluded as counter-models of $P = k \vee \neg c$.

5. Another option is to provide the pair $(K, \preceq_\Sigma^{lex})$ with a more refined ordering taking into account the number of violated formulas. The lexico pre-order (denoted $\preceq_\Sigma^{lex}$) can be obtained by first stratifying the formulas of K as follows: for each default rule of Δ we compute the degree of necessity of its propositional counterpart: $\neg t \vee c$ is violated by $\{\omega_3, \omega_4, \omega_{11}, \omega_{12}\}$, $\neg c \vee \neg r$ is violated by $\{\omega_1, \omega_5, \omega_9, \omega_{13}\}$, $c \vee r$ is violated by $\{\omega_4, \omega_8, \omega_{12}, \omega_{16}\}$ leading to assign to all these three formulas the same necessity, hence the same stratum. All formulas in P are in a separated stratum with the highest priority (i.e., with a necessity degree equal to 1). It allows us to refine the previous pre-order $\preceq_\Sigma^{bo}$ on interpretations by counting the violated formulas in each stratum. For instance ω_{11} and ω_{12} where equivalent with respect to the best-out ordering $\preceq_\Sigma^{bo}$. Considering lexico-ordering $\preceq_\Sigma^{lex}$ we get a tuple $(0, 1)$ for ω_{11} meaning that ω_{11} violates 0 formula in the stratum of highest priority and 1 in the other stratum (namely $\neg t \vee c$), we get $(0, 2)$ for ω_{12} (namely violating $\neg t \vee c$ and $c \vee r$) yielding $\omega_{11} \preceq_\Sigma^{lex} \omega_{12}$. The lexico-ordering gives more precisely: $\{\omega_2, \omega_6, \omega_7, \omega_{15}\} \preceq_\Sigma^{lex} \{\omega_1, \omega_3, \omega_5, \omega_8, \omega_{11}, \omega_{16}\} \preceq_\Sigma^{lex} \{\omega_4, \omega_{12}\}$, the set $\{\omega_9, \omega_{10}, \omega_{13}, \omega_{14}\}$ being again excluded as counter-models of P.

After showing how a commonsense knowledge base naturally induces a belief revision operator, we end this background section with a brief review of surprise modeling in the literature.

[8] $\preceq_\Sigma^{bo}$ stands for best-out pre-ordering.

3.4 From nonmonotonicity to surprises

Surprise arises from the occurrence of something considered practically impossible or unexpected. The first version of the possibility theory (Zadeh (1978), Dubois and Prade (1988)) proposed by the English economist Shackle (1961) was based on the notion of the degree of surprise associated with an event, which was in fact a degree of impossibility of this event, calculated from a possibility distribution reflecting the uncertain knowledge about the state of the world considered. A fact A is then all the more surprising that it is less coherent with what is imagined possible. Formally,

$$\texttt{surpriseDegree}(A) = 1 - \texttt{possibility}(A).$$

In AI, the question of surprises has been studied by Lorini and Castelfranchi (2007) who associate to surprise an incomprehension due to the difficulty to integrate the new perception, "she cannot believe it". The degree of surprise or astonishment is then associated with the probability of this new perception formalized in the *logic of probabilistic quantified beliefs* by Fagin and Halpern (1994). For these authors, the notion of surprise is first defined by a form of inconsistency with explicit expectations or with a priori knowledge. Then, once the surprise has arrived, the agent must revise his knowledge.

We now are in position to present our formalization of joke understanding in terms of belief revision.

4 Formalization of surprising and revealing statements

One of the first models proposed in linguistics for the analysis of jokes is that of Suls (1972). It is a two-phase model: the punchline which is unexpected, followed by a resolution (thought in terms of problem solving) which restores consistency (see also Shultz (1976)). Ritchie (1999, 2002) works on riddles such as "Why do birds fly south in winter? It's too far to walk". He identifies the presence of four entities in such riddles: a simple (most obvious) interpretation of the context, a hidden meaning of this context, a meaning of the punchline and an interpretation formed with integrating the punchline with the hidden meaning of the context. Let us note that in other riddles the unexpectedness could rely on double-entente: "Postmaster: Here's your five-cent stamp. Shopper: Do I have to stick it on myself? Postmaster: Nope. On the envelope."

In the approach by Dupin de Saint-Cyr and Prade (2020), belief change is used to define the surprise and thus to model its integration in the listener's mind like Lorini and Castelfranchi did. Indeed the punchline of the joke seen as a new piece of information triggers a revision of the beliefs whose result is surprising in the sense that it is inconsistent with what was initially believed.

Thus, we formalized the listener manipulation as a two-phase process (Dupin de Saint-Cyr and Prade (2020, 2022)):

- The first phase is a revision of the listener's beliefs with incomplete information intended to suggest a conclusion (which will turn out to be false).
- The second phase corresponds to the arrival of the punchline, which on the one hand surprises because it is incompatible with the previous provisional conclusion, which is translated in our model by an inconsistency. Moreover, the punchline is revealing, which we express by the fact that it logically explains the initial information.

Moreover, in this study we limit ourselves to the handling of the reasoning carried out by an agent listening to a joke. We leave aside the reasoning that the speaker may have done to develop the joke by adapting it to what he knows about the audience, not to mention the evaluation of whether or not to tell the joke in the situation.

4.1 Formalization of surprises in jokes

A joke is a progressive story. This means that when the listener receives α (the context of the joke) he or she is put in a situation of reasoning under incomplete information. This is indeed the teller's trick to lead

the listener to wrong conclusions from α in order to finally surprise him or her with the punchline β. Indeed the teller manipulates the belief states of the listener. This is made possible because the listener is ready to use default reasoning to make sense of α. Consider the extreme case of a fully psycho-rigid listener who would use only strict rules with all possible exceptions explicitly mentioned. He or she would not jump to any conclusion from α at all, and would wait for β for understanding what is going on. Thus this psycho rigid listener cannot be surprised (he or she is a monotonic thinker) and is not open to laughing. [9]

We now propose a formalization of jokes in propositional logic, i.e., we consider that the joke describes a situation α and that the punchline completes this description by an information β. The information α and β are propositions of $\mathcal{L}$. We are interested in the case where the joke is heard by a listener whose knowledge base is a set of propositional formulas, here assimilated to its conjunction denoted K. Moreover the listener is also characterized by the way $\circ$ she revises her beliefs according to a pre-order $\preceq_K$ as explained in the previous section.

In the following, a joke is seen as a statement that is a pair (<context>, <punchline>) with particular characteristics. We define a statement as follows:

Definition 4.1. A *statement* is a pair (α, β) formulas of $\mathcal{L}$.

The cognitive situation induced by a statement is described by the knowledge base K revised successively by the two elements of the statement. The punchline of a joke is considered *surprising* if the result of the revision of K by the description of the initial situation is contradictory to what we get after a revision by the punchline. Let us mention that in order to be surprise the listen should first be able to make an expectation from the context she or he heard. This is why the first condition to be surprised is that $K \circ \alpha$ is consistent.

Definition 4.2. A statement (α, β) is *surprising* for a listener associated with $(K, \circ)$ if

$$(K \circ \alpha) \text{ is consistent}^{10} \text{ and } (K \circ \alpha) \wedge (K \circ (\alpha \wedge \beta)) \models \bot$$

Note that the above definition excludes the possibility that the listener already knows the joke. Indeed, in the latter case, α and β would already be deducible from K, prohibiting any surprise. Besides, we do not require that $K \circ (\alpha \wedge \beta)$ be consistent, enabling nonsensical punchlines in Definition 4.2[11].

Example 3.2 (continued): *Now given the piece of information t under the integrity constraints P the listener equated with $(K, \preceq_K)$ will revise K by t by computing $Mod(K \circ^P t) = \min(Mod(t), \preceq_K)$ and obtain that the only model of the world is ω_2. Assuming that the new information is $t \wedge r$, three models correspond to $K \circ^P (t \wedge r)$ namely ω_1, ω_3 and ω_{11} where ω_2 no longer appear, acknowledging that $(t, t \wedge r)$ is surprising for the listener.*

4.2 Potentially funny jokes

To understand the joke, its logic must seem implacable once the punchline is revealed, so the punchline is both admissible ($K \circ \beta$ consistent) and explains the situation. In other words, if one had known β from the beginning, it would have explained α. This can be translated as follows:

Definition 4.3. Given a statement (α, β), its punchline β is *revealing* for $(K, \circ)$ if

$$(K \circ \beta) \text{ is consistent}^{12} \text{ and } K \circ \beta \models \alpha$$

Laughter can be viewed as a way of relieving the tension of cognitive dissonance caused by the inconsistency between what one expected to hear and the punchline of the story. This relief can only take place once the story is understood, with the punchline playing the role of a revelation.

[9] This is illustrated by the well-known story in which a mathematician and a physician are clearly reluctant to make a close world assumption in order to jump to a default conclusion in presence of incomplete information: A mathematician, a physicist and a biologist are on a train in Ireland. Through the window, they see a black sheep. How interesting, exclaimed the biologist, in Ireland, sheep are black! - You can't say that, replied the physicist. Certainly, there is at least one black sheep in Ireland. - Come, come, continues the mathematician, the only thing that can be said is that there is at least one sheep, at least one side of which is black!

[10] This condition is automatically ensured by the usual revision postulates (see 3.1) as soon as α is consistent.

[11] However, nonsensical punchlines will not be revealing in the sense of the next definition 4.3 due to postulates (KM5) and (KM6).

[12] This condition is automatically ensured by the revision postulates as soon as β is consistent, see Section 3.1.

Definition 4.4. A statement is *potentially funny* for a listener associated with $(K, \circ)$ if it is surprising and its punchline is revealing for this listener.

Figure 1 pictures a representation of a pre-order $\preceq_K$ where each layer is a set of interpretations at the same plausibility level (the most plausible layer being at the bottom of the figure). Models of α are in the dashed circle, and those of β are in the dotted circle. The models of $K \circ \alpha$ and $K \circ \beta$ (which here are equal to $K \circ (\alpha \wedge \beta)$) are materialized by a thick dashed segment and a thick dotted segment respectively. One can check that the two segments are distinct one being strictly less plausible than the other making clear that $Mod(K \circ \alpha) \cap Mod(K \circ (\alpha \wedge \beta)) = \emptyset$, i.e., (α, β) is surprising for the listener $(K, \circ)$. Moreover $K \circ \beta \models \alpha$ hence the punchline β is revealing for our listener.

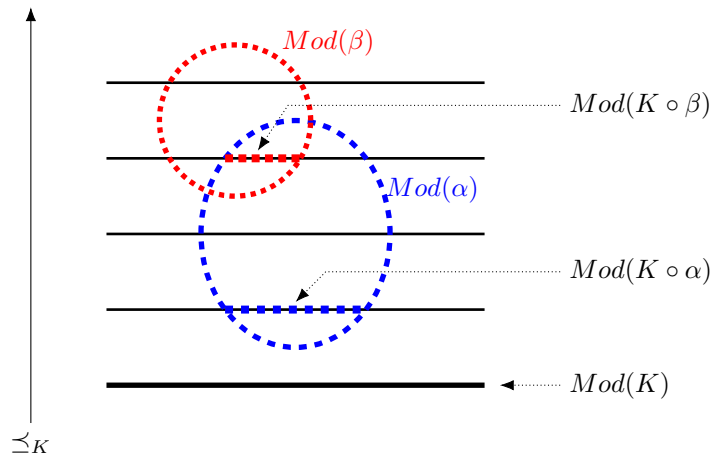

Figure 1: A typical situation in which the statement (α, β) is potentially funny

This definition of a potentially funny statement is directly inspired by Raccah (2015), except that we tried to define the revealing character of the punchline rather than its inevitability (which could be understood as $K \circ \alpha \models \beta$, denying the surprising character of β). Note that the definition relies on both K and $\circ$, which may differ from one listener to another, which is consistent with the fact that a joke may not be found funny by everyone.

The following proposition restates the definition of a potentially funny joke in a nonmonotonic setting: namely it shows how a listener equated with a pair $(K, \circ)$ induced from her or his defeasible knowledge base $\Sigma(P, \Delta)$, can find a joke surprising and revealing. Note that this definition is done for statements (α, β) consistent with P, in agreement with the conditions $K \circ \alpha$ and $K \circ \beta$ consistent of Definitions 3.4 and 4.3.

Proposition 4.5. *Let $\Sigma = (P, \Delta)$ be a knowledge base where P is a consistent set of propositional formulas and Δ is a consistent set of default rules and let $\preceq^{bo}_{\Delta}$ be the total best-out pre-order on the set of interpretations Ω uniquely defined from Δ. Let $(K, \circ)$ be such that $K = \{\neg\alpha_i \vee \beta_i | \alpha_i \rightsquigarrow \beta_i \in \Delta\}$ and $\circ$ be an operator defined from the total pre-order $\preceq^{bo}_{\Delta}$ for all $\varphi \in \mathcal{L}$ by: $Mod(K \circ \varphi) = \min(Mod(\varphi), \preceq^{bo}_{\Delta})$.*

Let $\alpha, \beta \in \mathcal{L}$ such that $\{\alpha, \beta\}$ is consistent with P,

- *The statement (α, β) is* surprising *for a listener, equated with $(K, \circ)$, if and only if* $\min(Mod(\alpha), \preceq^{bo}_{\Delta}) \cap \min(Mod(\alpha \wedge \beta), \preceq^{bo}_{\Delta}) = \emptyset$
- *The statement (α, β) is* revealing *for $(K, \circ)$ if and only if* $\beta \mid\!\sim^{bo}_{\Delta} \alpha$

where $\mid\!\sim^{bo}_{\Delta}$ is the non-monotonic inference relation defined by $\varphi \mid\!\sim^{bo}_{\Delta} \psi$ if and only if for each $\omega \in \min(Mod(\varphi), \preceq^{bo}_{\Delta})$, $\omega \models \psi$.

The same proposition holds for the lexicographic ordering, i.e., if we replace "bo" by "lex".

Proof. Let us first show that $\circ$ is a revision operator satisfying KM postulates: this is due to the fact that any interpretation that does not violate any default of Δ is in the set of the interpretations that are the most

preferred ones according to $\preceq^{bo}_{\Delta}$. Now, each interpretation that satisfies $K = \bigwedge_{\alpha_i \rightsquigarrow \beta_i \in \Delta} \neg\alpha_i \vee \beta_i$ does not violate any default, which means that the function that assigns to K the total pre-ordering $\preceq_K = \preceq^{bo}_{\Delta}$ is faithful, due to Theorem 3.1 we get that $\circ$ satisfies KM postulates.

Due to Definition 4.2, a statement (α, β) is surprising for $(K, \circ)$ if $(K \circ \alpha)$ is consistent and $(K \circ \alpha) \wedge (K \circ (\alpha \wedge \beta)) \models \bot$, due to the assumption that $\{\alpha, \beta\}$ is consistent with P, it means that α is consistent hence due to (KM3), $K \circ \alpha$ is consistent. Moreover by definition of $\circ$ based on Δ, $Mod(K \circ \alpha) = \min(Mod(\alpha), \preceq^{bo}_{\Delta})$ and $Mod(K \circ (\alpha \wedge \beta)) = \min(Mod(\alpha \wedge \beta), \preceq^{bo}_{\Delta})$. Hence the result.

Due to Definition 4.3, a statement (α, β) is revealing for $(K, \circ)$ if $K \circ \beta$ is consistent and $K \circ \beta \models \alpha$. Again, due to the assumption that $\{\alpha, \beta\}$ is consistent with P, it holds that β is consistent due to (KM3), $K \circ \beta$ is consistent. Moreover due to the definition of $\circ$ based on Δ, $Mod(K \circ \beta) = \min(Mod(\beta), \leq^{bo}_{\Delta})$. Hence the result.

This proof holds if we replace the best-out ordering with the lexicographic ordering, because the interpretations that do not violate any default are also the preferred ones for $\preceq^{lex}_{\Delta}$. □

We are now in position to show that a psycho-rigid listener (i.e., an agent that describes the world only with strict rules) cannot understand a potentially funny joke, because he cannot be surprised:

Proposition 4.6 (psycho-rigidity)**.** *If $\Sigma = (P, \Delta)$ is a knowledge base where P is a consistent set of propositional formulas and Δ is empty, then for any pair (α, β) such that $\{\alpha, \beta\}$ is consistent with P, the statement (α, β) is* not potentially funny *for the listener $(K, \circ)$ built from Σ as in the previous proposition.*

Proof. If Δ is empty then $\preceq^{bo}_{\Delta}$ ranks every interpretation at the same priority level, hence since α and $\alpha \wedge \beta$ are consistent, they have common minimum models with respect to $\preceq^{bo}_{\Delta}$. Thus, the statement (α, β) is not surprising for $(K, \circ)$. This holds also with the $\preceq^{lex}_{\Delta}$ pre-ordering. □

4.3 Joke examples and further propositions

We are now in a position to formalize a first example (from a French collection of "funny" stories by Nègre (1970)) expressed in natural language that we have transposed into logic. Note that in this article, we ignore the conversion stage from natural language to logic.

Example 4.7. A man has just been hit by a car. The driver gets out of the car and says: "You're lucky, we're right in front of a doctor's office. Yes! Except that the doctor is me!

Modeling:

$$\alpha = injured \wedge doctorNearby$$
$$\beta = injured \wedge doctorHimself$$

We assume that the listener has the following knowledge:

$$\Sigma = \left| \begin{array}{l} injured \wedge doctorNearby \rightsquigarrow treatedRapidly \\ injured \wedge \neg treatingDoctor \rightsquigarrow \neg treatedRapidly \\ doctorHimself \rightsquigarrow doctorNearby \\ injured \wedge doctorHimself \rightsquigarrow \neg treatingDoctor \end{array} \right.$$

As shown in Section 3.3, from $\Sigma = (P, \Delta)$, K and $\preceq_K$ can be built. The reader can check that we have the following pre-order on interpretations satisfying α:
$\{R\ \neg H\ T\ i\ N\} \preceq_K \{R\ H\ \neg T\ i\ N,\ R\ \neg H\ \neg T\ i\ N,\ \neg R\ H\ \neg T\ i\ N,\ \neg R\ \neg H\ T\ i\ N,\ \neg R\ \neg H\ \neg T\ i\ N\}$
$\preceq_K \{R\ H\ T\ i\ N,\ \neg R\ H\ T\ i\ N\}$ where R, H, T, i, N are short for $treatedRapidly$, $doctorHimself$, $treatingDoctor$, $injured$, $doctorNearby$ respectively.

Due to postulate (KM1), $K \circ \alpha \models \alpha$. More precisely, the only model of $K \circ \alpha$ is $R\ \neg H\ T\ i\ N$ (it is the most plausible model of α according to $\preceq_K$).
The models of $K \circ (\alpha \wedge \beta)$ are $R\ H\ \neg T\ i\ N$ and $\neg R\ H\ \neg T\ i\ N$. The punchline is surprising since $Mod(K \circ \alpha) \cap Mod(K \circ (\alpha \wedge \beta)) = \emptyset$.

Moreover, the punchline explains α: $K \circ \beta \models i \wedge N$ since $Mod(K \circ \beta)$ is the set of the best models of β according to $\preceq_K$, which turn out to be here the same two models as the ones obtained for $K \circ (\alpha \wedge \beta)$. Indeed, $N = doctorNearby$ is the part of the joke that manipulated us by leading us to think that the wounded man was lucky.

Note that Example 4.7 can also be seen as a cascading joke (φ_1, φ_2, φ_3, φ_4) with $\varphi_1 = injured$, $\varphi_2 = lucky$, $\varphi_3 = doctorNearby$, $\varphi_4 = doctorHimself$. Indeed, saying "You are very lucky" after having noticed that the person is hurt can already provoke laughter. Then, the revelation of being "in front of a doctor's office" helps to understand. Finally, the second part of the joke corresponds to what has already been analyzed. Note that in the first part of the cascading joke, φ_2 is surprising but not revealing, it is φ_3 that explains φ_2. So for a cascading joke, the surprising and revealing effects are not necessarily simultaneous.

The following property shows the importance of narrative order in the joke.

Proposition 4.8 (Never give the punchline before the end!)**.** *If (α, β) is a potentially funny statement for a listener knowing K, then (β, α) is not potentially funny for that listener.*

Proof. Since (α, β) is potentially funny, then the punchline is revealing, that is, $K \circ \beta \not\models \bot$ and $K \circ \beta \models \alpha$. But that prevents the story (β, α) from being surprising, and therefore potentially funny. Indeed, using (KM5) and (KM6), $K \circ (\beta \wedge \alpha) \equiv (K \circ \beta) \cup \{\alpha\}$ so $K \circ \beta$ is consistent with $K \circ (\beta \wedge \alpha)$. □

The next property indicates the virtue of brevity.

Proposition 4.9 (the shortest jokes...)**.** *If (α, β) is a potentially funny statement knowing K then adding $\alpha' \in \mathcal{L}$, such that $\alpha \not\models \alpha'$, can make the statement $(\alpha \wedge \alpha', \beta)$ not potentially funny.*

Proof. Indeed here is a counterexample: let $\Sigma = (P, \Delta)$ where $P = p \rightarrow b$, $\Delta = \{b \rightsquigarrow f, p \rightsquigarrow \neg f, b \wedge a \rightsquigarrow p\}$ a knowledge base expressing that: generally birds (b) fly (f), penguins (p) are birds, generally penguins (p) don't fly, most Antarctic birds (a) are penguins. Let us consider $(K, \circ)$ built as explained in Section 3.3. Let us imagine a story (b, p), it is both surprising since $K \circ b \models b \wedge \neg p \wedge \neg a \wedge f$ and $K \circ (b \wedge p) \models b \wedge p \wedge \neg f$ and the "punchline" p is revealing since $K \circ p \models b$. However, by adding information $\alpha' = a$ that the bird is from Antarctica we can create an unsurprising situation with respect to $\beta = p$, so $(b \wedge a, p)$ is no longer a surprising story. Another counterexample consists in adding information α' that is not related to β then we can obtain a punchline that is not completely revealing: let us consider $\alpha' = t$, where t means "is called Tweety" then $K \circ p \models b$ does not allow us to deduce that $K \circ p \models t$. □

To convince ourselves of the validity of the Property 4.9, we can consider the particular case $\alpha' = \beta$. This would correspond in the example 4.7 to specifying that the driver has just run over a doctor in front of the doctor's office. In this case, the punch line "Yes! except that the doctor is me! is no longer funny (because we already know that). Nevertheless the new story can be found funny in a different way with $\alpha'' =$ *A man has just run over his own doctor in front of his office* and $\beta'' =$ *You're very lucky, we're right in front of my doctor's office*. Here, the statement is quite surprising, but the revelation can only come if we perceive the irony of the use of the word lucky (in place of unlucky).

We can notice that 4.9 does not require that the punchline be as short as possible. The following definition refines the definition of "funny" into "efficient" by imposing minimality (in terms of interpretations).

Definition 4.10 (conciseness of the punchline)**.** Let $\alpha, \beta, \beta' \in \mathcal{L}$ be individually consistent formulas and $\beta \not\equiv \beta'$, if $K \circ \beta \models \beta'$ then the statement (α, β) is said to be more *efficient* than the statement (α, β') for listener knowing K.

In other words, the punchline is said to be more efficient if it is more precise and therefore more specific. Here is another example from Nègre (1970)'s collection.

Example 4.11. A guy and a woman enter the doctor's office. The doctor turns to the lady and says :
• If you are ill, please undress...
But the girl has manners. She is reluctant. She looks at the guy from underneath. So the doctor repeats:
• But come on, lady, get undressed! I am a doctor. There's no indecency, here! Then she starts to fidget and suddenly bursts into tears. Disconcerted, the doctor asks the guy:
• What's wrong with your wife? Is she always so nervous?
• I don't know. I just met her in your waiting room...

Modeling:

$\varphi_1 = together, \quad \varphi_2 = reluctant, \quad \varphi_3 = together \wedge \neg knowEachOther$

$$\Sigma = \left| \begin{array}{ll} P & \neg knowEachOther \rightarrow \neg couple \\ \hline & together \rightsquigarrow couple \\ \Delta & couple \rightsquigarrow \neg reluctant \\ & \neg couple \rightsquigarrow reluctant \end{array} \right.$$

As computed in Example 3.2 where k, t, c, r are short for $knowEachOther$, $together$, $couple$, $reluctant$, the only model of $K \circ \varphi_1$ is ω_2, the models of $K \circ (\varphi_1 \wedge \varphi_2)$ are ω_1 and ω_3 hence φ_2 is surprising since $Mod(K \circ \varphi_1) \cap Mod(K \circ (\varphi_1 \wedge \varphi_2)) = \emptyset$. Moreover in the lexicographic version of the pre-ordering over interpretations ($\preceq_\Sigma^{lex}$) the punchline φ_3 explains φ_2: since the most plausible model of φ_3 according to $\preceq_\Sigma^{lex}$ is ω_{11} hence $K \circ \varphi_3 \models reluctant$. Note that this result is not obtained with the best-out pre-ordering $\preceq_\Sigma^{bo}$ which is not able to discriminate between ω_{11} and ω_{12} as noticed in Example 3.2.

Suppose now that the guy replied φ_3' =“Sorry, we entered together but we are not married” ($\varphi_3' = together \wedge \neg couple$). It is clear that $K \circ \varphi_3 \models \varphi_3'$. Hence the old punchline φ_3 is more efficient than the new punchline φ_3'.

Note that we might have considered iterated revision Konieczny and Pino Pérez (2000) and abductive expansion Pagnucco (1996). However, in order to capture the cognitive aspects of understanding a joke, the main idea is to select the best epistemic state rather than to maintain the order of preferences over all states as in iterated revision. Indeed, in the basic revision, the epistemic states can be radically changed. The surprise would not be the same if the auditor had in mind all the potential explanations. In this respect, a similar argument applies to abductive expansion, whose goal is to provide reasons for the new epistemic state. Furthermore, we have favored the simplicity of classical revision theory.

4.4 Graduality in surprising and revealing jokes

A lesson that we draw from Raccah’s article concerns the intensity of funniness: according to Raccah (2015), “the joke is all the funnier that the trap was unexpected and inevitable”. This leads us to propose to define two quantities to classify the jokes. This would allow us to build a partial order on the jokes addressed to the same listener reflecting their potential funniness. This definition assumes that the knowledge of the listener contains default rules, as previously assumed, it should have the form: $\Sigma = (P, \Delta)$ enabling us to compute a possibility distribution π_Σ. The surprise level is conform to the view recalled in subsection 3.4. The second term is the conditional necessity of α knowing $K \circ \beta$, reflecting the ineluctability of the revelation.

Definition 4.12 (gradual effects in jokes)**.** Given a base $\Sigma = (\Delta, P)$, a statement (α, β) is associated to the possibilistic levels:

- a surprise level: $\begin{cases} 0 & \text{if } N(\alpha) = N(\alpha \wedge \beta), \\ 1 - \Pi(\alpha \wedge \beta) & \text{otherwise} \end{cases}$
- a revealing level: $N(\alpha | K \circ \beta)$

where $K \circ \varphi$ is defined from Σ by building $(K, \preceq_K)$ and N and Π are the necessity and possibility measures associated with π_Σ, as explained in subsection 3.2.

Let’s not forget that other elements can contribute to the funniness, as the huge literature on humor can testify. All these additional aspects are themselves gradual in nature: such as the cognitive effort necessary to understand the joke, the language level, the comic effects of the narrator, the mood of the listener, his capacity of inhibition of his emotions, etc. This is a topic for further research.

Moreover the modeling of laughter is a multi-agent problem involving a teller and a listener, especially when we come to the handling of social notions like complicity between listener and teller, and acceptable incongruity: the teller should take into account what he knows of the listener’s norms. We now further discuss incongruity.

5 Integrating incongruity

This section deals with inconsistency (associated with incongruity) as a new feature in jokes; this time inconsistency occurs with respect to listener's (and teller's) knowledge and not because of the listener's ability to reason with incomplete information. We first discuss incongruity and then absurdity as a particular case, before introducing the idea of "acting as if" a law / norm could be ignored, which is related to incongruity as we shall see.

5.1 Introducing incongruity and absurdity

As said in Section 2, incongruity is usually related to the violation of social norms (i.e., lack of conformity with current practices, unsuitable behavior). Let us start with examples.

Example 5.1.
• To drink water from a finger bowl is an *incongruous* behavior because it violates hygienic norms.
• Wearing oversized shoes is *incongruous* because shoes are made for easy walking. This is why seeing a clown with such shoes has a comic effect all the more as it causes his falling.

One can notice there that the comic effect occurs because the joke is deliberately in disregard with an obvious universal knowledge. Acting as if a universal law was not mandatory is an incongruous behavior.

As illustrated by the next example, a suitable behavior when communicating with people is to respect common knowledge about the object / situation under discussion. When someone derogates to this social principle, an *absurd* communication takes place.

Example 5.2.
• Speaking about someone drinking the water from a finger bowl is a description of an *incongruous* behavior but it is *not absurd* with respect to communication: it is not impossible that someone does it.
• Speaking about carrying an elephant in an ordinary car (with 4 seats) is an *absurdity*, because as everybody knows this is not feasible (an elephant is too big and too heavy).

So absurdity may be viewed as a special form of incongruity as far as it comes from the lack of respect of some communication principle. Note that incongruity, hence absurdity, may be felt as surprising inasmuch incongruity is unusual (since resulting from the violation of current practices).

One can observe that in many comic situations, some "acting as if" is at work. This means that a person behaves without taking into account some obvious facts or pieces of knowledge about the world. Thus the clown strives to walk normally despite of his unsuitable shoes. In the four elephants riddle, one strives to reason normally despite of the feasible impossibility of carrying elephants in a car, as discussed now.

Example 5.3.
i, e, tt, h being propositional formulas representing respectively the fact that 4 elephants are in the 2CV car, four elephants are in presence, two elephants are in the front and two elephants are in the rear of the 2CV car, huge animals are in presence.
$\alpha : i \wedge e$
β: $tt \wedge e$
$\Sigma = (P, \Delta)$ with $P = \{R1, R2, R3, R4\}$ and $\Delta = \emptyset$ where
R1:$h \rightarrow \neg i$ (Something huge cannot be inside a 2CV car)
R2: $e \rightarrow h$ (Elephants are huge)
R3: $h \rightarrow \neg tt$ (Something huge cannot be put in the rear and in the front)
R4: $tt \rightarrow i$ (Putting in the rear and in the front implies putting inside)

$P \cup \{\alpha\}$ is inconsistent, hence the notion of surprising for a listener (whose revision operator is induced by Σ) given in previous section (Proposition 4.5) does not apply. Another kind of surprise is at stake: the violation of an obvious law (R2: elephants are huge), thus we are in presence of an absurdity in communication. In this example the absurdity appears already in α: $P \models \neg\alpha$. But the revealing part is still there, when forgetting this obvious law (R2), namely, $\{\beta\} \cup (P \setminus \{R2\}) \models \alpha$ hence β is revealing for an agent knowing Σ and ignoring $R2$.

We now give another example where incongruity comes from the violation of social norms.

Example 5.4. My house caught fire (fh), my neighbor welcomes me to his house. I set fire to his house (sFn). You wonder why. Because he told me to do what I do at home ((an): allowed to make himself at home at the neighbor's), and my home was in fire.
α = fh $\wedge$ sFn
β = fh $\wedge$ an
$\Sigma = (P, \Delta)$ with $P = \{R2\}$ and $\Delta = \{R1, R3, R4\}$ where
R1: sFn $\rightsquigarrow$ ir (set fire at the neighbor's is generally an irrational act)
R2: fh $\wedge$ dh $\rightarrow$ sFn (having a fire at home and doing like at home implies to setting fire at neighbor's)
R3: an $\rightsquigarrow \neg$ ir (when allowed to make yourself at home generally you do not perform irrational acts)
R4: an $\rightsquigarrow$ dh (if you are allowed to make yourself at home then generally you do like at home)

The use of "allowed" is a way to encode permission, here there is a violation of a social permission (rule $R3$), the storyteller is not allowed to destroy the house of his neighbor despite being allowed to make himself at home. Nevertheless there is a revealing part, because with β and $(P, \Delta \setminus \{R3\})$ we get α (we reason logically in disregard of $R3$ which encodes an expected basic attitude).

Note that, in the previous example, the surprise comes from the fact that an obvious law is violated, the revealing side of the joke is obtained by ignoring this particular piece of knowledge. It leads us to define a new notion that was already implicitly present in the examples formalized in the previous section: the violation of a social norm, and the revealing with disregard of a formula.

5.2 Formal handling of incongruity

Before defining formally incongruity, we introduce the nonmonotonic entailment operator used for representing the listener reasoning. In the following definition, a knowledge base $\Sigma = (P, \Delta)$ induces a pre-order $\preceq^m_\Sigma$ on the set of interpretations Ω (m being either the best-out or the lexicographic method) where$\preceq^m_\Sigma$ is the restriction of $\preceq^m_\Delta$ to the models of P. We define $\mid\!\sim^m_\Sigma$, the nonomonotonic inference relation based on Σ, as follows. In this definition, P acts as an integrity constraint, hence entailment is possible only from formulas consistent with P.

Definition 5.5 (nonmonotonic entailment under $\Sigma = (P, \Delta)$)**.** $\varphi \mid\!\sim^m_\Sigma \psi$ if
- $P \cup \{\varphi\}$ is consistent and
- for all $\omega \in \min(Mod(P \cup \{\varphi\}) \preceq^m_\Sigma), \omega \models \psi$.

The notion of disregard is now defined in the setting of a knowledge base Σ composed with strict and default rules from which non-monotonic reasoning can be performed, as described in subsection 3.3. For capturing this notion we would like to express that the reasoning should lead to conclude that a formula representing a norm is violated.This means that the whole joke (the context of the punch line) contradicts at least one norm: this is why the following definition characterizes the disregard of one formula ρ, considered as a norm by the listener. In presence of such an incongruity the revealing in the sense of Definition 4.3 is no more desirable since we would like to reason while ignoring the norm. This leads us to the following definition.

Definition 5.6 (incongruity)**.** Given $\Sigma = (P, \Delta)$, let $P_N \subseteq P$ and $\Delta_N \subseteq \Delta$ be subsets of formulas considered as norms in Σ, let ρ be a formula either in P_N or in Δ_N, the statement (α, β) is *felt incongruous* by a listener represented by Σ if:

1. $\alpha \wedge \beta \mid\!\sim^m_{\Sigma\setminus\{\rho\}}$ Not ρ (violation of ρ)

2. $\beta \mid\!\sim^m_{\Sigma\setminus\{\rho\}} \alpha$ (revelation in disregard of ρ)

where, when $\rho \in P$, $\Sigma \setminus \{\rho\} = (P \setminus \{\rho\}, \Delta)$ and Not $\rho = \neg\rho$;
when $\rho = \alpha_i \rightsquigarrow \beta_i \in \Delta$, $\Sigma \setminus \{\rho\} = (P, \Delta \setminus \{\rho\})$ and Not $\rho = \alpha_i \wedge \neg\beta_i$.

Note that a particular case of violation (1) is when $\alpha \wedge \beta$ is inconsistent with P. Besides, some jokes cannot be felt incongruous because they violate an integrity constraint that cannot be forgotten, as made clear by the following proposition.

Proposition 5.7. *Any formula $\rho \in P$ such that $P \setminus \{\rho\} \models \rho$ cannot be considered as violated by a listener (P, Δ).*

Proof. By monotonicity of classical logical consequence, we would have $\{\alpha \wedge \beta\} \cup P \setminus \{\rho\} \models \rho$. Definition 5.5 would lead to have for all $\omega \in \min(Mod(\{\alpha \wedge \beta\} \cup P \setminus \{\rho\}), \omega \models \neg\rho$ which is contradictory. Hence the result. □

The situation is different when $\rho = \alpha_k \rightsquigarrow \beta_k \in \Delta$, it is possible that $\alpha \wedge \beta \mid\!\sim^m_{\Sigma \setminus \{\rho\}} \alpha_i \wedge \neg\beta_i$ holds together with $\alpha_i \mid\!\sim^m_{\Sigma \setminus \{\rho\}} \beta_i$ (i.e., the knowledge base $\Sigma \setminus \{\rho\}$ still entailing the default ρ). However, it is not a problem to allow such formulas to be disregarded because in presence of $\alpha \wedge \beta$ the nonmonotonic mechanism will favor the inference of $\alpha_i \wedge \neg\beta_i$, hence the violation of the norm ρ.

In other words, incongruity is not so much a matter of *removing* a norm from the listener knowledge base but rather a *violation* of this norm by focusing on exceptional situations. For instance in the story about the four elephants, if we consider that all information is defeasible, then the violation of the rule that elephants are too huge to enter in a car is possible by considering exceptional elephants with no huge size. But such a consideration is by itself incongruous by violating the general law.

We should notice that Definition 5.6 is using a set difference operator in order to check if by removing ρ in the knowledge base Σ, the statement $\alpha \wedge \beta$ violates ρ and also to check if β reveals α when removing ρ. The reader might wonder if $\Sigma \setminus \{\rho\}$ should be better replaced by a contraction[13], however, when $\rho \in \Delta$, the removal of ρ is in fact not mandatory for checking incongruity, since the only thing that we would like to know is whether ρ is violated in the context $\alpha \wedge \beta$. When $\rho \in P$, for showing that ρ is violated we should at least remove ρ from P if we want to reason safely (without inconsistency).

The removal of $\rho \in P$ by mere erasing of ρ from P, may have some side effects due to proposition 5.7 which forbids a listener to consider formulas that are entailed by other formulas of P as violated in the sense of the first condition of Definition 5.6. It appears that when this side effect takes place this could be understood as the fact that the listener is reluctant to the intended incongruity of the joke because she or he is not able to apprehend all the inconsistency it creates. When this effect does not occur the listener is able to handle the violation thus to appreciate the joke. This is the case in the two examples below where the norm violated cannot be derived from the rest of the knowledge base. The first one does not involve any default rule and illustrates the violation of a norm represented by a propositional formula. The second one concerns the violation of a norm expressed by a default rule.

Example 5.3 (continued): *In this example, $\alpha = i \wedge e$, $\beta = e \wedge tt$, Δ is empty, and R2 is taken as the only norm, hence $\varphi \mid\!\sim^m_{\Sigma} \psi$ reduces to $P \cup \{\varphi\} \models \psi$ when $P \cup \{\varphi\}$ is consistent. We have $\{i \wedge e \wedge tt\} \cup P \setminus \{R2\} \models e \wedge \neg h$. Moreover, $P \setminus \{R2\} = \{R1, R3, R4\}$ is consistent with $i \wedge e \wedge tt$ and there is only one model satisfying them namely $\omega_0 = i\ e\ tt\ \neg h$ in which $e \wedge \neg h$ holds, negating $\rho = R2 = e \rightarrow h$. Hence the statement $(i \wedge e, tt \wedge e)$ is violating $R2$.*

Now, the set of models of $tt \wedge e$ satisfying $(P \setminus \{R2\})$ reduces to $\{\omega_0\}$, and $\omega_0 \models i \wedge e$, hence the statement $(i \wedge e, tt \wedge e)$ is a revelation in disregard of $R2$ for a listener corresponding to Σ.

Example 5.4 (continued): *In this example, R1 and R3 are taken as norms, $\alpha = fh \wedge sFn$, $\beta = fh \wedge an$. We are going to show that $fh \wedge sFn \wedge an \mid\!\sim^{bo}_{\Sigma \setminus \{R3\}} an \wedge ir$. Indeed after removing the rule R3, the rules R1 and R4 have the same priority level in Δ. Note that among the interpretations satisfying $fh \wedge sFn \wedge an$, the interpretation $\omega_1 = fh\ an\ ir\ dh\ sFn$ is the only one that does not violate any rule, hence it is the only preferred one. $\omega_1 \models an \wedge ir$, thus violating $\rho = R3 = an \rightsquigarrow \neg\ ir$.*

[13] We recall that the contraction of a knowledge base by φ is the operation that modifies it in order that it no longer entails φ. This operation was introduced in a propositional logic setting by Alchourrón et al. (1985). More formally, K contracted by ρ, denoted K$\dot{-}$ ρ, can be computed as a revision using $(K \circ \neg\rho) \wedge K$ according to Harper's identity (Gärdenfors (1988)). However, such a modeling does not seem very satisfactory from a cognitive point of view. Indeed a listener willing to understand a joke, is not necessarily inclined to consider all the consequences of the violation of a norm. It is enough for a listener to find one norm that is violated, independently of the existence of sets of formulas that entail this norm. Moreover, as far as we know the contraction by default rules has not been much studied.

When removing $R3$ the only preferred model is ω_1 which satisfies sFn, hence fh $\wedge$ an$\mid\!\sim^{bo}_{\Sigma\setminus\{R3\}}$ fh $\wedge$ sFn which means that (fh $\wedge$ sFn, fh $\wedge$ an) is a revelation in disregard of $R3$.

As already suggested in introducing incongruity as a supplementary ingredient of humor, the two examples of subsection 4.3 that where only considered as potentially funny (since being surprising and revealing), may also be seen as involving incongruity. Indeed, in example 4.7 we could underline the incongruity of the situation (an injured doctor) by the fact that usually doctors are always ready to cure people, which could be encoded by a rule such as $doctorHimself \rightsquigarrow \neg injured$. In example 4.11, there is an implicit social principle that is disregarded, namely "a man should not follow a lady that he does not know (especially in a doctor's office)". More precisely, this rule could be encoded as $\neg knowEachOther \rightsquigarrow \neg together$ which is violated by β.

Example 4.11 (continued): *If we add the norm $\rho = \neg k \rightsquigarrow \neg t$ to Δ then the statement (t, r) is still surprising, then the statement $(r, t \wedge \neg k)$ is violating ρ and it is a revelation in disregard of ρ. Thus the story $(t, r, t \wedge \neg k)$ is felt both incongruous and surprising by a listener represented by $(P, \Delta \cup \{\rho\})$.*

As we are going to see in the following remark, using the archetypal example of nonmonotonic reasonning, an exception to a default rule could be understood as incongruous to the extent that this default rule can be regarded as a norm. This is natural since when a new situation is presented by a teller, the listener is first inclined towards considering the situation as normal.

Remark 1. Given a knowledge base $\Sigma = (P, \Delta)$ where $P = \{p \rightarrow b\}$ and $\Delta = \{R1 = b \rightsquigarrow f, R2 = p \rightsquigarrow \neg f, R3 = b \wedge a \rightsquigarrow p\}$, the statement (b, p) is potentially funny (as shown in Proof of Proposition 4.9) and it is incongruous because a typical rule R1 is violated: $p \wedge b \mid\!\sim^{bo}_{\Sigma\setminus\{R1\}} b \wedge \neg f$. Indeed we are in a case of an exception to rule R1.

In this subsection, we have provided a first attempt at modeling incongruity in jokes. We have used the setting introduced in Section 4 where the listener is identified with a a pair (P, Δ) acknowledging her / his capability of reasoning under incomplete knowledge, leading us to depart from a pure abstract belief revision framework. This proposal remains preliminary, and there may be different ways to extend this setting. For instance, we may think about introducing degrees of incongruity, we could also take into account the mixed emotions of the listener such as the shame associated to the violation of norms. Let us also admit that identifying what is a norm may depend from a listener to another this echoes the fact that the same joke cannot be understood in the same way by people with different background.

5.3 Discussion on the relevance of the proposed model

Without any doubt, a large number of jokes rely on a surprising punchline in the sense that the teller presents an incomplete situation that puts the listener on a wrong track at the beginning of the joke, before the punchline reveals the reality of the situation. The fact that such a mechanism is present in many jokes is acknowledged by authors specialized in the study of jokes such as in Ritchie (2018) who himself outlined a modeling of the kind discussed in Section 4, although in a much less formal way (Ritchie (1999, 2002)).

Nevertheless, it is true that there are circumstances other than jokes where we can be surprised at the end, as in detective stories for instance, which are not especially funny in general. Indeed in this paper we speak of "potentially funny" jokes, since on the one hand not all listeners laugh at the same jokes, and on the other hand the triggering of laughter also depends in part on the teller. Laughter is a matter of cognitive science which also involves physiological, psychological, sociological and cultural dimensions.

Still, surprise and revelation remain ingredients that are at work in many jokes as noticed by philosophers and psychologists. But there are other kinds of surprising situations that are not based on incomplete information at the beginning, but rather rely on inconsistency, as in the elephants joke. Then, to escape inconsistency, one has to "act as if" inconsistency was not there (a known recipe for comical effects), we then speak of incongruity. Incongruity can be reinforced if the inconsistency involves a socially forbidden situation, for example, setting the neighbor's house on fire, or undressing in front of strangers.

It is difficult to determine what characterizes the specificity of what leads to laughter, and over the centuries philosophers have only been able to identify a few ingredients, the three main ones being the surprise effect, the incongruity and the superiority. Indeed all the jokes considered here exhibit some incongruity and revelation (and possibly some surprise due to incomplete information), even if there exist

jokes which may be also based on hazardous analogies (see Dupin de Saint-Cyr and Prade (2022) for some examples).

Whatever the case (surprising punchline or incongruity) the joke requires a resolution, this is its revealing side. This modest contribution has proposed a formalization of the two ingredients. Interestingly enough, they are directly related to two basic issues in artificial intelligence: incomplete information and inconsistency.

6 Conclusion

In this article, dedicated to Philippe Besnard, we have revisited our proposal to see jokes as a matter of surprise and revelation expressed in terms of belief revision. The key ideas of this paper are i) the claim that the capability of the listener of reasoning under incomplete information is necessary for understanding a joke and ii) the notion of incongruity can be integrated in a nonmontonic reasoning framework. Incongruity is seen as a violation of a social norm, the comic effect comes from the fact that the story teller acts as if this norm is not existing.

This preliminary proposal could still be enriched in different ways. First, one may consider that the comic effect is all the greater as the disregarded social norm is more obvious or mandatory. This means that one may introduce some level of incongruity in relation with the shocking nature of the non-compliance with the involved law.

As recalled in the introduction, another acknowledged ingredient of humor is superiority. Taking this notion into account would require a more sophisticated framework for representing the communication process between agents that could reason and speak about other agents. This calls for a genuine multiple agent modeling. Such a framework would allow for representing how a teller can take into account what he or she knows about the listener and how the teller can create a sense of complicity with the listener, possibly against a common object of mockery. In such a context, principles of good communication (such as Grice (1975)'s maxims) can be at stake. This allows for jokes that are incongruous with respect to such principles (like absurd jokes that disregard the principle of always taking the reality into account).

In general, understanding a story is facilitated by the ability to visualize and put oneself in the place of the characters. This helps to perceive the blatant side of the violation of the law. Imagining a character in a story capable of breaking an obvious law may allow the listener for a kind of "comic release" (from the pressure to follow that law). At the opposite, a story in a universe with people you can hardly identify with has little chance of making you laugh. However, in this context, it is possible to achieve laughter by helping the listener to identify with the characters. In Valitutti and Veale (2016), geometric forms are moving on a screen where a small circle seems to follow a big one. This scene is shown to an audience that do not laugh at it. The same scene is shown with a fictional dialog between the shapes involving a man and a woman, then it provokes the laughter of the audience. Lastly, a story itself may be more or less funny, but the way of telling it is also important (it may involve incongruity, since the storyteller might disregard some Gricean-like maxims).

In the same way as surprise and revelation are not enough for uniquely characterizing jokes, it is not sure that the addition of incongruity would be sufficient for capturing the comic effect of the joke and enough for characterizing it in a unique way. Indeed Spencer (1860) emphasized that incongruity is not necessarily a source of laughter since it can also provoke wonder, as sometimes in art.

Acknowledgements

This paper has greatly benefited from the remarks of the reviewers.

References

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: partial meet contraction and revision functions. *The Journal of Symbolic Logic*, *50*(2), 510–530.

Attardo, S., & Raskin, V. (1991). Script theory revis(it)ed: joke similarity and joke representation model. *Humor-International Journal of Humor Research*, *4*(3-4), 293–348.

Bariaud, F. (1983). *La génèse de l'humour chez l'enfant*. Presses Universitaires de France, Paris.

Beattie, J. (1776). An essay on laughter, and ludicrous composition. *Essays. William Creech, Edinburgh. Reprinted by Garland, New-York 1971*, 414–456.

Benferhat, S., Cayrol, C., Dubois, D., Lang, J., & Prade, H. (1993). Inconsistency management and prioritized syntax-based entailment. In R. Bajcsy (Ed.), *Proc. 13th int. joint conf. on artificial intelligence. Chambéry, Aug. 28 - Sept. 3* (pp. 640–647). Morgan Kaufmann.

Benferhat, S., Dubois, D., & Prade, H. (1992). Representing default rules in possibilistic logic. In B. Nebel, C. Rich, & W. R. Swartout (Eds.), *Proc. 3rd int. conf. on principles of knowledge representation and reasoning (KR'92). Cambridge, MA, Oct. 25-29* (pp. 673–684). Morgan Kaufmann.

Benferhat, S., Dubois, D., & Prade, H. (1997). Nonmonotonic reasoning, conditional objects and possibility theory. *Artif. Intell.*, *92*(1-2), 259–276.

Benferhat, S., Dubois, D., Prade, H., & Williams, M.-A. (2002). A practical approach to revising prioritized knowledge bases. *Studia Logica*, *70*(1), 105–130.

Bergson, H. (1900). *Le rire: Essai sur la signification du comique*. Felix Alcan, Paris.

Chapman, A. J., & Foot, H. C. (Eds.). (1976). *Humour and laughter: Theory, research and applications*. John Wiley & Sons.

Delage, Y. (1919). Sur la nature du comique. *La Revue du Mois*, *20*, 337-354.

Descartes, R. (1649). *Les passions de l'âme*. Henry Le Gras, Paris.

Dictionary, M.-W. (2023). *Incongruous.* Retrieved from `https://www.merriam-webster.com/dictionary/incongruous`

Dubois, D., & Prade, H. (1988). *Possibility theory. an approach to computerized processing of uncertainty*. New York and London: Plenum Press. (with the collaboration of H. Farreny, R. Martin-Clouaire and C. Testemale)

Dubois, D., & Prade, H. (1991). Epistemic entrenchment and possibilistic logic. *Artif. Intell.*, *50*(2), 223–239.

Dupin de Saint-Cyr, F., & Prade, H. (2020). Jokes and belief revision. In D. Calvanese, E. Erdem, & M. Thielscher (Eds.), *Proc. 17th int. conf. on principles of knowledge representation and reasoning (KR'20), rhodes, sept. 12-18* (pp. 336–340).

Dupin de Saint-Cyr, F., & Prade, H. (2022). La compréhension des histoires drôles : une affaire de révision de croyances. *Revue Ouverte d'Intelligence Artificielle (ROIA)*, *3*, 253–285.

Fagin, R., & Halpern, J. Y. (1994). Reasoning about knowledge and probability. *J. of the ACM (JACM)*, *41*(2), 340–367.

Fleet, F. R. (1890). *A theory of wit and humour*. Remington and Company.

Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states*. MIT Press, 1988. (2nd ed., College Publications, 2008)

Gärdenfors, P. (1990). Belief revision and nonmonotonic logic: Two sides of the same coin? In *Proceedings of the 9th european conference on artificial intelligence (ECAI'90), Stockholm* (pp. 768–773).

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics. vol. 3: Speech acts* (p. 41—58). Academic Press, New York.

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic*, 157–170.

Hobbes, T. (1651). *Leviathan*. Crooke, London Press.

Hollingworth, H. L. (1911). Experimental studies in judgment: Judgments of the comic. *Psychological Review*, *18*(2), 132.

Kant, E. (2001). *Critique of the power of judgment*. Cambridge University Press. (Original title: Kritik der Urteilskraft, Lagarde und Friederich, Berlin u. Libau, 1790)

Katsuno, H., & Mendelzon, A. O. (1991). Propositional knowledge base revision and minimal change. *Artificial Intelligence*, *52*(3), 263–294.

Katsuno, H., & Mendelzon, A. O. (1992). On the difference between updating a knowledge base and revising it. In P. Gärdenfors (Ed.), *Belief revision* (p. 183-203). Cambridge University Press.

Keith-Spiegel, P. (1972). Early conceptions of humor: Varieties and issues. In J. H. Goldstein (Ed.), *The psychology of humor: Theoretical perspectives and empirical issues* (p. 3-39). Academic Press New York.

Konieczny, S., & Pino Pérez, R. (2000). A framework for iterated revision. *J. Applied Non-Class. Logics*, *10*, 339–367.

Latour, M. (1956). *Le problème du rire et du réel. quatre essais*. Presses Universitaires De France, Paris.

Léa Sombé (research group). (1994). A glance at revision and updating in knowledge bases. *Int. J. of Intelligent Systems*, *9*(1), 1–27.

Lehmann, D. (1995). Belief revision, revised. In *Proc. 14th int. joint conf. on artificial intelligence-volume 2* (pp. 1534–1540).

Lorini, E., & Castelfranchi, C. (2007). The cognitive structure of surprise: looking for basic principles. *Topoi*, *26*(1), 133–149.

Maier, N. R. F. (1932). A gestalt theory of humour. *British Journal of Psychology*, *23*(1), 69-74.

Nègre, H. (1970). *Dictionnaire des histoires drôles*. Fayard.

Pagnol, M. (1947). *Notes sur le rire*. Editions Nagel, Paris.

Pagnucco, M. (1996). *The role of abductive reasoning within the process of belief revision*. Univ. Sydney Dissertation.

Pearl, J. (1990). System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proc. 3rd conf. on theoretical aspects of reasoning about knowledge* (pp. 121–135).

Raccah, P.-Y. (2015). Humour et métaphore: quelques éléments d'une analogie pour la construction d'un sens inattendu: Illustration sur un corpus de citations de George Bernard Shaw. *Revue de Sémantique et Pragmatique, Presses de l'Univ. d'Orléans*. Retrieved from `<halshs-01099484>`

Ritchie, G. (1999). Developing the incongruity-resolution theory. In *Proc. of the AISB symposium on creative language: Stories and humour, Edinburgh* (p. 78–85).

Ritchie, G. (2002). The structure of forced reinterpretation jokes. In O. Stock, C. Strapparava, & A. Nijholt (Eds.), *Proc. 20th Twente workshop on language technology (TWLT20): The April Fools' Day workshop on computation humour, Trento, Apr. 15-16* (p. 47-56).

Ritchie, G. (2018). *The comprehension of jokes: A cognitive science framework*. CRC Press.

Schopenhauer, A. (1819). *The world as will and representation*. English edition 1969, Peter Smith Publisher. (Original title: Die Welt als Wille und Vorstellung)

Sen, A. K. (1971). Choice functions and revealed preference. *The Review of Economic Studies*, *38*(3), 307–317.

Shackle, G. L. S. (1961). *Decision, order and time in human affairs*. UK: (2nd edition), Cambridge University Press.

Shultz, T. R. (1976). A cognitive-developmental analysis of humour. In A. J. Chapman & H. C. Foot (Eds.), *Humour and laughter: Theory, research and applications* (pp. 11–36). John Wiley & Sons.

Spencer, H. (1860). The physiology of laughter. *Macmillan's Magazine*, *1*, 395-402.

Suls, J. M. (1972). A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. *The psychology of humor: Theoretical perspectives and empirical issues*, *1*, 81–100.

Thea Arbee. (2016, March 5). *Is the superfluous a (new) modality?* (Talk given at IRIT, Toulouse, at the occasion of a meeting celebrating the Emeritus title of Luis Fariñas del Cerro. Unfortunately unpublished)

Valitutti, A., & Veale, T. (2016). Infusing humor in unexpected events. In *Proc. international conference on distributed, ambient, and pervasive interactions* (pp. 370–379).

Winslett, M. (1988). Reasoning about action using a possible models approach. In H. E. Shrobe, T. M. Mitchell, & R. G. Smith (Eds.), *Proc. of the 7th national conference on artificial intelligence, St. Paul, Aug. 21-26, 1988* (pp. 89–93). AAAI Press / The MIT Press.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, *1*, 3–28.