Fuzzy Autoepistemic Logic: Reflecting about Knowledge of Truth Degrees

Marjon Blondeel[1] ⋆, Steven Schockaert[2] ⋆⋆, Martine De Cock[2], and Dirk Vermeir[1]

[1] Dept. of Computer Science, Vrije Universiteit Brussel, Belgium
{mblondee,dvermeir}@vub.ac.be
[2] Dept. of Applied Mathematics and Computer Science, Ghent University, Belgium
{steven.schockaert,martine.decock}@ugent.be

Abstract. Autoepistemic logic is one of the principal formalisms for nonmonotonic reasoning. It extends propositional logic by offering the ability to reason about an agent's (lack of) knowledge or beliefs. Moreover, it is well known to generalize the stable model semantics of answer set programming. Fuzzy logics on the other hand are multi-valued logics, which allow to model the intensity with which a property is satisfied. We combine these ideas to a fuzzy autoepistemic logic which can be used to reason about one's knowledge about the degrees to which proporties are satisfied. In this paper we show that many properties from classical autoepistemic logic remain valid under this generalization and that the important relation between autoepistemic logic and answer set programming is preserved in the sense that fuzzy autoepistemic logic generalizes fuzzy answer set programming.

1 Introduction

Autoepistemic logic was introduced by Moore [16] as a way to reason about one's own beliefs. Later on (e.g. [15]), it was also seen as a tool to reflect about one's (lack of) knowledge. Consider for example my reason for believing that my sister smokes. If she smoked, I would have smelled it on her breath and I would believe she smoked: "smoke → breath" and "breath → B(smoke)", where B means "I believe". Now suppose I have never smelled anything, thus I do not believe that she smokes, then I can conclude that she does not smoke.

Since its introduction in the 1980s, autoepistemic logic has been one of the principal formalisms for nonmonotonic reasoning. It has also found important applications in logic programming. For example, Gelfond and Lifschitz [7] showed a connection between answer sets of logic programs and expansions of autoepistemic theories.

Fuzzy logics (e.g. [8]) are a class of logics, whose semantics are based on truth degrees that are taken from the unit interval $[0, 1]$. By admitting intermediary

⋆ Funded by a joint Research Foundation-Flanders (FWO) project
⋆⋆ Postdoctoral fellow of the Research Foundation-Flanders (FWO)

truth values between 0 and 1, the intensity with which some property holds can be encoded. From a practical point of view, fuzzy logics are thus useful to model knowledge of continuous domains in a logical way. For example, instead of saying that my sister smokes or not, a value between 0 and 1 can be given to specify how much she smokes. Reconsider the rule "smoke $\rightarrow$ breath", but in the setting of fuzzy logics. Then we may interpret this rule as "If she smoked a lot, her breath would smell often."

In this paper we combine the ideas of autoepistemic logic and fuzzy logics, which to the best of our knowledge, has not previously been considered. The resulting fuzzy autoepistemic logic is useful to reflect on one's beliefs (or knowledge) about the *degrees* to which some properties are satisfied. Consider for example my reason for not believing that my sister smokes *a lot*. If she smoked *a lot*, her breath would smell *often*. Since I do not smell it *often*, I do not believe she smokes *a lot*. Intuitively, if the truth value of $\mathrm{B}\varphi$ is equal to c, this means that we only know that φ is true *at least* to degree c. Note in particular that the degrees of belief which we consider do not reflect strength of belief, but rather a Boolean form of belief in graded properties. Furthermore, note how this view generalizes the notion of belief from classical autoepistemic logic, in the sense that having $\mathrm{B}\varphi$ false corresponds to having φ true to at least degree 0 and having $\mathrm{B}\varphi$ true corresponds to having φ true at least to degree 1.

In this paper we show that many important proporties from classical autoepistemic logic remain valid when generalizing to fuzzy autoepistemic logic. We also prove that the relation between autoepistemic logic and answer set programming is preserved. In particular we show that the answer sets of a fuzzy answer set program correspond to the stable expansions of an associated fuzzy autoepistemic logic theory. The fact that this important relationship is preserved provides further insight into the nature of fuzzy answer set programming, and at the same time serves as a justification for the particular fuzzy autoepistemic logic we introduce in this paper.

The paper is structured as follows. In Section 2 the necessary background on autoepistemic logic and fuzzy logic is given. In Section 3 fuzzy autoepistemic logic is introduced, its properties are investigated and a motivating example is given. In Section 4, we briefly recall the basic notions of a fuzzy version of answer set programming which was recently proposed and we analyze the relation with fuzzy autoepistemic logic. We finish the paper by discussing related work and our conclusions in Sections 5 and 6.

2 Background

2.1 Autoepistemic Logic

The formulas of autoepistemic logic are built from a set of propositional atoms A using the usual propositional connectives and a modal operator B, interpreted as "is believed" (or "is known"). For example, if φ is a formula, then $\mathrm{B}\varphi$ indicates that φ is believed. Also, $\mathrm{B}(\neg\varphi)$ indicates that $\neg\varphi$ is believed and $\neg(\mathrm{B}\varphi)$ that φ

is not believed. We write L for the language of all propositional formulas over A, and L_{B} for the extension of L with the modal operator B. As done in the literature (e.g. [13]), formulas from L are called *objective* and formulas from L_{B} are called *unimodal*. An *autoepistemic theory* is a set of unimodal formulas. We define $A' = A \cup \{\mathrm{B}\varphi \mid \varphi \in L_{\mathrm{B}}\}$, which is an infinite set, even if A is finite. For technical reasons, we sometimes treat A' as a set of atoms, and consider interpretations $I' \in \mathcal{P}(A')$, where $\mathcal{P}(X) = \{Y \mid Y \subseteq X\}$ for a set X. This trick allows us to deal with autoepistemic theories in a purely propositional fashion. For clarity, we will refer to the corresponding propositional language as L'.

The following definition was introduced by Moore [16] and is in line with Stalnaker's [20][3] view on the beliefs of a rational agent. For an arbitrary autoepistemic theory T, we can look for maximally conservative extensions which make it stable, in the sense that no more conclusions can be drawn from a stable theory than what is explicitly contained in it. Such extensions are called *stable expansions* of T.

Definition 1. *Suppose E and T are autoepistemic theories, then E is a* stable expansion *of T iff*

$$E = \mathrm{Cn}(T \cup \{\mathrm{B}\varphi \mid \varphi \in E\} \cup \{\neg\mathrm{B}\varphi \mid \varphi \notin E\}),$$

where $\mathrm{Cn}(X)$ *denotes the set of propositional consequences of X w.r.t. the language L'.*

Remark that Definition 1 says that a formula α is in E iff for each interpretation $I' \in \mathcal{P}(A')$ such that $I' \models T \cup \{\mathrm{B}\varphi \mid \varphi \in E\} \cup \{\neg\mathrm{B}\varphi \mid \varphi \notin E\}$ we have that $I' \models \alpha$. Moreover, the set of models of $T \cup \{\mathrm{B}\varphi \mid \varphi \in E\} \cup \{\neg\mathrm{B}\varphi \mid \varphi \notin E\}$ is exactly the set of models of E. By using Definition 1, the following proposition can be proved.

Proposition 1. *[14] Suppose T is a consistent autoepistemic theory. If all formulas in T are objective, then T has exactly one stable expansion.*

Autoepistemic logic can also be described in terms of models, more like a possible worlds semantics [17]. The relationship between this semantics and the one used in Definition 1 will become clear in Proposition 2. Suppose $I \in \mathcal{P}(A)$ is an interpretation on A and $S \subseteq \mathcal{P}(A)$ a set of interpretations on A. The corresponding satisfaction relation for unimodal formulas is defined inductively:

- For an atom p, $(I, S) \models p$ iff $p \in I$.
- For a unimodal formula φ, $(I, S) \models \mathrm{B}\varphi$ iff for every $J \in S$, $(J, S) \models \varphi$.
- For unimodal formulas φ and ψ, the propositional connectives are handled in the usual way:
  - $(I, S) \models (\varphi \wedge \psi)$ iff $(I, S) \models \varphi$ and $(I, S) \models \psi$
  - $(I, S) \models (\varphi \vee \psi)$ iff $(I, S) \models \varphi$ or $(I, S) \models \psi$
  - $(I, S) \models (\neg\varphi)$ iff $(I, S) \not\models \varphi$

[3] Article based on the unpublished manuscript (1980) to which Moore referred.

- $(I, S) \models (\varphi \rightarrow \psi)$ iff $(I, S) \not\models \varphi$ or $(I, S) \models \psi$
- $(I, S) \models (\varphi \leftrightarrow \psi)$ iff $(I, S) \models (\varphi \rightarrow \psi)$ and $(I, S) \models (\psi \rightarrow \varphi)$

Intuitively, a unimodal formula φ is *believed* to be true, if it is true in every interpretation which is considered possible.

Definition 2. *Suppose T is an autoepistemic theory and S is a set of interpretations on A, then S is an* autoepistemic model *of T iff*

$$S = \{I \mid I \in \mathcal{P}(A), \forall \varphi \in T : (I, S) \models \varphi\} .$$

In other words, an autoepistemic model is a set of interpretations which model all formulas of T.

Definition 3. *Suppose S is a set of interpretations on A and T is an autoepistemic theory, then T is called the* (autoepistemic) theory *of S iff*

$$T = \{\varphi \mid \varphi \in L_{\mathrm{B}}, \forall I \in S : (I, S) \models \varphi\} ,$$

We will write $\mathrm{Th}(S)$ *to denote this set of formulas.*

The set $\mathrm{Th}(S)$ contains all formulas that are true in every interpretation in S. The following proposition describes the relation between stable expansions and autoepistemic models.

Proposition 2. *[17] Suppose T is an autoepistemic theory, then an autoepistemic theory E is a stable expansion of T iff $E = \mathrm{Th}(S)$ for some autoepistemic model S of T.*

We will now discuss the relationship between answer set programming [7] and autoepistemic logic. A brief refresher on answer set programming is provided in Appendix A. Gelfond and Lifschitz [7] proposed the following transformation from a program P to an autoepistemic theory $\lambda(P)$. For each rule $s \leftarrow a_1, \ldots, a_m, \text{not } b_1, \ldots, \text{not } b_n$ in P, the unimodal formula $a_1 \wedge \ldots \wedge a_m \wedge \neg \mathrm{B} b_1 \wedge \ldots \wedge \neg \mathrm{B} b_n \rightarrow s$ is added to $\lambda(P)$. The following result clarifies the relationship between the answer sets of P and the stable expansions of $\lambda(P)$.

Theorem 1. *[6],[7] A logic program P has an answer set[4] M iff $\lambda(P)$ has a stable expansion E such that $M = E \cap \mathcal{B}_P$.*

2.2 Fuzzy Logics

Fuzzy logics [8] are based on an infinite number of truth degrees, taken from the unit interval $[0, 1]$. We will consider fuzzy logics whose formulas are built from a set of atoms A, the truth constants in $[0, 1] \cap \mathbb{Q}$ and arbitrary n-ary

[4] We refer to Appendix A for definitions and notations regarding answer set programming.

connectives for each $n \in \mathbb{N}$. In particular, the semantics of logical conjunction can be generalized to $[0,1]$ by a class of functions called *triangular norms* (short *t-norms*). These are mappings $\mathcal{T} : [0,1]^2 \to [0,1]$ which are symmetric, associative and increasing and which satisfy $\mathcal{T}(1,x) = x$ for all $x \in [0,1]$. Given a t-norm $\mathcal{T}$, logical implication can be generalized by the *residuation* of $\mathcal{T}$, $x \to_{\mathcal{T}} y = \sup\{\lambda \mid \lambda \in [0,1] \text{ and } \mathcal{T}(x,\lambda) \leq y\}$. If $\mathcal{T}$ is a left-continuous t-norm we have the important property $x \to_{\mathcal{T}} y = 1$ iff $x \leq y$. By using the residuation it is possible to define a generalization of the logical equivalence $x \leftrightarrow_{\mathcal{T}} y = \min\{x \to_{\mathcal{T}} y, y \to_{\mathcal{T}} x\}$. Negation can be generalized by a decreasing map $\sim : [0,1] \to [0,1]$ satisfying $\sim 1 = 0$ and $\sim 0 = 1$. In what follows we will only use residual implicators based on left-continuous t-norms and if there is no confusion possible we will write $x \to y$ and $x \leftrightarrow y$.

An interpretation is a mapping $I : A \to [0,1]$, which is also called a *fuzzy set* on A. We can extend this interpretation as follows. Consider for each $n \in \mathbb{N}$ a finite set of n-ary connectives F_n and let $F = \cup F_n$. Each $f \in F_n$ is interpreted by a function $\mathtt{f} : [0,1]^n \to [0,1]$. We define $[f(\alpha_1, \ldots, \alpha_n)]_I = \mathtt{f}([\alpha_1]_I, \ldots, [\alpha_n]_I)$ for formulas α_i $(1 \leq i \leq n)$. For $c \in [0,1]$ we have $[c]_I = c$. If T is a set of formulas we say that I is a model of T iff $[\alpha]_I = 1$ for all $\alpha \in T$; we write this as $I \models T$.

In examples we will consider the connectives from Łukasiewicz logic, however all theorems can be proved for connectives $f \in F$. In the case of Łukasiewicz logic, the conjunction is defined as $x \otimes y = \max(x + y - 1, 0)$, which is a left-continuous t-norm. The disjunction is generalized by $x \oplus y = \min(x + y, 1)$. The implicator induced by the Łukasiewicz t-norm is $x \to_l y = \min(1, 1 - x + y)$ and for the negation we have $\neg x = 1 - x$.

3 Fuzzy Autoepistemic Logic

In this section, we combine the ideas of autoepistemic logic and fuzzy logics. This will provide us a tool to reason about one's beliefs about the degrees to which one or more properties are satisfied. Let us consider an example, for which we will use Łukasiewicz logic. Note that the main results in this section are valid for arbitrary connectives in F.

Example 1. Suppose we want to host a party for three persons. Since we do not know how much each guest will eat, it is not easy to determine how much food we need to order. Let us denote this latter amount by a variable a that ranges between ordering no food at all and ordering the maximum amount of food. Obviously, the correct value for a depends on the amount of food a_i $(i = 1,2,3)$ that we need to order for each individual guest. The variable a_i represents the proposition that person i eats a full portion. For an interpretation I, $I(a_i)$ denotes which percentage of a full portion person i eats. By appropriately rescaling the food quantities we can assume, without lack of generality, that $I(a_i) \in [0, \frac{1}{3}]$ for each graded interpretation I, such that it holds that $I \models a_1 \oplus a_2 \oplus a_3 \leftrightarrow_l a$

iff $I(a_1) + I(a_2) + I(a_3) = I(a)$. We thus consider the following formulas.

$$a_i \to_l \frac{1}{3} \tag{1}$$

$$a_1 \oplus a_2 \oplus a_3 \leftrightarrow_l a. \tag{2}$$

If no further information about the values a_i is known, it is best to make sure that everybody has enough food by ordering the maximum amount. By encoding additional beliefs we will try to refine this upper bound. Suppose we believe that everyone will eat at least a certain amount of food. We express this as

$$\mathrm{B}a_1 \leftrightarrow_l 0.1, \mathrm{B}a_2 \leftrightarrow_l 0.1, \mathrm{B}a_3 \leftrightarrow_l 0.05. \tag{3}$$

As in classical autoepistemic logic, we can treat formulas $\mathrm{B}\varphi$ as atoms. For each interpretation I, we then have that $I \models \mathrm{B}a_i \leftrightarrow_l c_i$ iff $[\mathrm{B}a_i]_I = c_i$. Later on, it will become more clear why including a formula such as $\mathrm{B}a_i \leftrightarrow_l c_i$ expresses a lower bound for the truth value of a_i. For each model I we have that $[\mathrm{B}a_i]_I = c_i$ implies $I(a_i) \geq c_i$.

Furthermore, we assume that if someone would eat an exceptional amount of food, we would have some information about this. For example, this could be the case if our friend brings her new boyfriend. If he would have an extreme appetite, we believe that she would have warned us. Insisting that we would know exactly how much each person would eat, i.e. $a_i \to_l \mathrm{B}a_i$, would be too strong. We may consider the following weaker variant however, which expresses that no guest will eat more than three times the amount mentioned in (3). We represent this meta-knowledge as follows:

$$a_i \to_l \mathrm{B}a_i \oplus \mathrm{B}a_i \oplus \mathrm{B}a_i. \tag{4}$$

Indeed, $I \models a_i \to_l \mathrm{B}a_i \oplus \mathrm{B}a_i \oplus \mathrm{B}a_i$ iff $I(a_i) \leq 3[\mathrm{B}a_i]_I$. In addition, we may be able to further decrease the amount of food that needs to be ordered if we know that some of the guests are on a diet. We will represent this by a variable d_i which represents the proposition that person i is on an extreme diet. If $I(d_i) = 0$, person i eats like he/she normally eats and if $I(d_i) = 1$, he/she will eat the amount mentioned in (3). Suppose we have information on d_2 and d_3, but no knowledge on d_1:

$$\mathrm{B}d_2 \leftrightarrow_l 0.95, \mathrm{B}d_3 \leftrightarrow_l 0.95. \tag{5}$$

If the lower bound for d_i increases, the upper bound for a_i should decrease. Consider for instance the meta-knowledge

$$\mathrm{B}d_i \to_l (a_i \to_l \mathrm{B}a_i). \tag{6}$$

Remark that this expression is equivalent to $a_i \to_l (\mathrm{B}d_i \to_l \mathrm{B}a_i)$, thus $I \models \mathrm{B}d_i \to_l (a_i \to_l \mathrm{B}a_i)$ iff $I(a_i) \leq 1 + [\mathrm{B}a_i]_I - [\mathrm{B}d_i]_I$.

In example 4, we will use fuzzy autoepistemic logic to determine an upper bound for $I(a)$ for a model I.

The formulas in fuzzy autoepistemic logic are built from a set A (atoms and constants in $[0,1] \cap \mathbb{Q}$), the set of connectives F with their corresponding functions $\mathtt{f} : [0,1]^n \to [0,1]$ ($n \in \mathbb{N}$) and a modal operator B, interpreted as "is believed". We will denote this language as $\overline{L_\mathrm{B}}$. Again, we will make the distinction between objective and unimodal formulas. An *autoepistemic theory in* $\overline{L_\mathrm{B}}$ is a set of formulas in $\overline{L_\mathrm{B}}$. As before, we define $A' = A \cup \left\{\mathrm{B}\varphi \mid \varphi \in \overline{L_\mathrm{B}}\right\}$. We write $\mathcal{F}(A')$ for the set of all fuzzy sets on A', i.e. the set of all graded interpretations I' over A'. We define a generalization of stable expansions (Definition 1).

Definition 4. *Suppose T is an autoepistemic theory in $\overline{L_\mathrm{B}}$ and E is a fuzzy set on $\overline{L_\mathrm{B}}$. E is a* fuzzy stable expansion *of T iff for each $\alpha \in \overline{L_\mathrm{B}}$*

$$E(\alpha) = \inf\left\{[\alpha]_{I'} \mid I' \models T \cup \left\{\mathrm{B}\varphi \leftrightarrow E(\varphi) \mid \varphi \in \overline{L_\mathrm{B}}\right\}, I' \in \mathcal{F}(A')\right\}$$

In Definition 1, for each $I' \in \mathcal{P}(A')$ such that $I' \models T \cup \{\mathrm{B}\varphi \mid \varphi \in E\} \cup \{\neg\mathrm{B}\varphi \mid \varphi \notin E\}$ we had that $\mathrm{B}\varphi \in I'$ iff $\varphi \in E$. To see the relation between Definitions 1 and 4, note that I' is a model of $\mathrm{B}\varphi \leftrightarrow E(\varphi)$ iff $[\mathrm{B}\varphi]_{I'} = E(\varphi)$.

Remark that for a fuzzy stable expansion E of T and $I', J' \in \{I' \mid I' \models T \cup \{\mathrm{B}\varphi \leftrightarrow E(\varphi) \mid \varphi \in L_\mathrm{B}\}, I' \in \mathcal{F}(A')\}$ we have that $[\mathrm{B}\alpha]_{I'} = E(\alpha) \leq [\alpha]_{J'}$. Hence, $[\mathrm{B}\alpha]_{I'} = 0.1$ intuively means that α is true to at least degree 0.1, instead of believing that α is true to exactly degree 0.1.

We can also generalize Definitions 2 and 3. First, we need to define another type of evaluation for unimodal formulas. Suppose $I \in \mathcal{F}(A)$ is an interpretation and $S \subseteq \mathcal{F}(A)$ is a set of interpretations.

- For an atom or a constant p: $[p]_{I,S} = I(p)$.
- For a unimodal formula α: $[\mathrm{B}\alpha]_{I,S} = \inf_{J\in S}[\alpha]_{J,S}$.
- For unimodal formulas α_i ($1 \leq i \leq n$) and $f \in F_n$: $[f(\alpha_1,\ldots,\alpha_n)]_{I,S} = \mathtt{f}([\alpha_1]_{I,S},\ldots,[\alpha_n]_{I,S})$.

Definition 5. *Suppose T is an autoepistemic theory in $\overline{L_\mathrm{B}}$ and $S \subseteq \mathcal{F}(A)$ is a set of interpretations. S is a* fuzzy autoepistemic model *of T iff*

$$S = \{I \mid I \in \mathcal{F}(A), \forall\varphi \in T : [\varphi]_{I,S} = 1\}.$$

Example 2. Suppose $T = \{\neg(\mathrm{B}a) \to_l b, \neg(\mathrm{B}b) \to_l a\}$. We try to find a fuzzy autoepistemic model S of T. For the first formula of T we have for $S \subseteq \mathcal{F}(A)$ and $I \in S$ that $[\neg(\mathrm{B}a) \to_l b]_{I,S} = 1 \Leftrightarrow 1 - [\mathrm{B}a]_{I,S} \leq I(b) \Leftrightarrow 1 - I(b) \leq \inf_{J\in S} J(a)$. By symmetry we have $[\neg(\mathrm{B}b) \to_l a]_{I,S} = 1 \Leftrightarrow 1 - I(a) \leq \inf_{J\in S} J(b)$.

Hence, a set of interpretations S is a fuzzy autoepistemic model of T iff $S = \{I \mid I \in \mathcal{F}(A), 1 - I(b) \leq \inf_{J\in S} J(a) \text{ and } 1 - I(a) \leq \inf_{J\in S} J(b)\}$. Moreover, we can show that the fuzzy autoepistemic models of T are all sets of the form $S_x = \{I \mid I \in \mathcal{F}(A), I(a) \geq x \text{ and } I(b) \geq 1 - x\}$, with $x \in [0,1]$.

Definition 6. *Suppose $S \subseteq \mathcal{F}(A)$ is a set of interpretations. The* fuzzy autoepistemic theory *of S is the fuzzy set $\mathrm{Th}(S)$ on $\overline{L_\mathrm{B}}$ such that for each unimodal formula φ*

$$\mathrm{Th}(S)(\varphi) = \inf_{I\in S}[\varphi]_{I,S}.$$

We can prove the following generalizations of Propositions 1 and 2.

Proposition 3. *Suppose T is an autoepistemic theory in $\overline{L_{\mathrm{B}}}$. A fuzzy set E on $\overline{L_{\mathrm{B}}}$ is a fuzzy stable expansion of T iff $E = \mathrm{Th}(S)$ with S a fuzzy autoepistemic model of T.*

Example 3. Reconsider the theory T from Example 2 and recall that all fuzzy autoepistemic models are of the form $S_x = \{I \mid I(a) \geq x \text{ and } I(b) \geq 1 - x\}$. Hence, for each $x \in [0,1]$ we have a fuzzy stable expansion E_x defined by $E_x(a) = \mathrm{Th}(S_x)(a) = \inf_{I \in S_x} I(a) = x$ and $E_x(b) = \mathrm{Th}(S_x)(b) = \inf_{I \in S_x} I(b) = 1 - x$.

Proposition 4. *Suppose T is a consistent set of objective formulas in $\overline{L_{\mathrm{B}}}$, then there is exactly one fuzzy set E on $\overline{L_{\mathrm{B}}}$ that is a fuzzy stable expansion of T.*

Example 4. Reconsider Example 1. Based on the formulas (2)-(6), an upper bound for $I(a)$ (I a model) can be derived. This is accomplished by determining the fuzzy autoepistemic models of the corresponding autoepistemic theory T. Suppose $S \subseteq \mathcal{F}(A)$, then we determine which conditions need to be satisfied for $I \in S$ such that S is a fuzzy autoepistemic model of T:

$$\begin{gathered}
I(a_i) \leq \tfrac{1}{3} \\
I(a_1) + I(a_2) + I(a_3) = I(a) \\
[\mathrm{B}a_1]_{I,S} = 0.1,\ [\mathrm{B}a_2]_{I,S} = 0.1,\ [\mathrm{B}a_3]_{I,S} = 0.05 \\
I(a_i) \leq 3[\mathrm{B}a_i]_{I,S} \\
[\mathrm{B}d_2]_{I,S} = 0.95,\ [\mathrm{B}d_3]_{I,S} = 0.95 \\
[\mathrm{B}d_i]_{I,S} \leq 1 - I(a_i) + [\mathrm{B}a_i]_{I,S}
\end{gathered}$$

For example, let us compute the upper bound for $I(a_2)$. Without the knowledge about the diet, we know that $I(a_2) \leq 0.3$. If we include our beliefs about d_2, we get a much lower upper bound 0.15.

One can easily verify that there is exactly one fuzzy autoepistemic model

$$\begin{aligned}
S = \{I \mid\ & 0.1 \leq I(a_1) \leq 0.3, 0.1 \leq I(a_2) \leq 0.15, 0.05 \leq I(a_3) \leq 0.10, \\
& I(d_1) \geq 0, I(d_2) \geq 0.95, I(d_3) \geq 0.95, 0.25 \leq I(a) \leq 0.55\}\,.
\end{aligned}$$

We thus believe that the amount of food that will be needed is between 0.25 and 0.55. Hence we will order 55% of the maximal order. Note that this means that we can express the lower bound on a as $E(\mathrm{B}a) = 0.25$ and the upper bound as $E(\neg\mathrm{B}(\neg a)) = 0.55$, where $E = \mathrm{Th}(S)$ is the unique stable expansion of T.

4 Relation between Fuzzy Answer Set Programming and Fuzzy Autoepistemic Logic

Let us briefly recall the basic notion of a fuzzy version of answer set programming, which was recently proposed [10]. Consider a set of atoms A. Here, a *literal* is either an atom $a \in A$ or an expression of the form not a, where $a \in A$ and not is the negation-as-failure operator. A rule over $[0,1]$ is an expression of the form

$r : a \leftarrow f(b_1, \ldots, b_n)$ where $a \in A$, b_i $(1 \leq i \leq n)$ are literals, $\leftarrow$ corresponds to a residual implicator and $f \in F_n$. To assure the existence of a unique answer set we need to restrict to connectives f such that $\mathtt{f}$ is increasing in each argument. Typically $\mathtt{f}$ corresponds to the application of conjunctions and disjunctions in a given fuzzy logic. We will refer to the rule by its label r. The atom a is called the head of r and $f(b_1, \ldots, b_n)$ is the body. A *FASP program* over $[0,1]$ is a set of rules over $[0,1]$. We denote the set of atoms occurring in a FASP program as $\mathcal{B}_P$. An *interpretation* I of a FASP program P is a mapping $I : \mathcal{B}_P \rightarrow [0,1]$. We can extend this mapping as follows:

- $[c]_I = c$ for $c \in [0,1]$,
- $[\text{not } a]_I = \sim([a]_I)$ for atoms a and a negator $\sim$,
- $[f(b_1, \ldots b_n)]_I = \mathtt{f}([b_1]_I, \ldots [b_n]_I)$ for bodies of rules,
- $[r]_I = ([r_b]_I \rightarrow I(r_h))$, for a rule $r : r_h \leftarrow r_b$.

For interpretations I_1 and I_2 we say that $I_1 \leq I_2$ iff $I_1(a) \leq I_2(a)$ for all $a \in \mathcal{B}_P$. An interpretation I is called a *model* of P iff $[r]_I = 1$ for all $r \in P$. Finally we say that a FASP program is *simple* if it contains no literals of the form not a. For such programs there exists a unique minimal model.

Definition 7. *[10] Consider a simple FASP program P. An interpretation I of P is called the* answer set *of P iff it is the minimal model of P.*

For programs which are not simple, answer sets are defined using a generalization of the Gelfond-Lifschitz reduct (see Appendix A). Specifically, let P be a FASP program and I an interpretation of P. The *reduct* of a literal l w.r.t. I is defined as follows. If l is an atom then $l^I = l$, if $l = \text{not } a$ then $l^I = [l]_I$. The reduct of a rule in P, $r : a \leftarrow f(b_1, \ldots, b_n)$ is defined as $r^I : a^I \leftarrow f(b_1^I, \ldots, b_n^I)$. The reduct of the program P is the set of rules $P^I = \left\{r^I \mid r \in P\right\}$.

Definition 8. *[10] Consider a FASP program P. An interpretation I of P is called an answer set of P iff I is the answer set of P^I.*

In this section we will show a correspondence between answer sets of a FASP program P and fuzzy stable expansions of an associated autoepistemic theory in $\overline{L_{\mathrm{B}}}$. From Theorem 1, we already know that such a correspondence exists between classical ASP and autoepistemic logic. Here we use a similar transformation. Suppose we have a FASP program P with rules of the form

$$r : a \leftarrow f(b_1, \ldots, b_n, \text{not } c_1, \ldots, \text{not } c_m),$$

where a, b_i and c_j are atoms $(1 \leq i \leq n)$,$(1 \leq j \leq m)$ and $f \in F_{n+m}$. We define a set of implications in fuzzy autoepistemic logic. Specifically, for rule r we define the associated fuzzy autoepistemic formula $\lambda(r)$ as

$$f(b_1, \ldots, b_n, \sim_1 \mathrm{B}c_1, \ldots, \sim_m \mathrm{B}c_m) \rightarrow a.$$

We choose $\sim_i$ as the negator which is assumed for not c_i, thus for $I \in \mathcal{F}(\mathcal{B}_P)$, we have $[\text{not } c_i]_I = \sim_i (I(c_i))$. The resulting autoepistemic theory in $\overline{L_{\mathrm{B}}}$ is $\lambda(P) = \{\lambda(r) \mid r \in P\}$.

First, we provide a lemma that characterizes the relationship between stable expansions of $\lambda(P)$ and stable expansions of the autoepistemic theory corresponding to a specific reduct of the program P. Note that we use the notation $E_{|\mathcal{B}_P}$ for the restriction of the fuzzy set E on $\overline{L_{\mathrm{B}}}$ to $\mathcal{B}_P$.

Lemma 1. *Consider a FASP program P and a fuzzy set E on $\overline{L_{\mathrm{B}}}$. Then E is a fuzzy stable expansion of $\lambda(P)$ iff E is a stable expansion of $\lambda(P^{\bar{E}})$, where $\bar{E} = E_{|\mathcal{B}_P}$.*

Theorem 2. *Consider a FASP program P. M is an answer set of P iff $\lambda(P)$ has a fuzzy stable expansion E such that $E_{|\mathcal{B}_P} = M$.*

Example 5. Consider the logic program $P = \{b \leftarrow_l \text{not } a, a \leftarrow_l \text{not } b\}$. We will compute the answer sets by using the characterization from Theorem 2. We look for fuzzy stable expansions of $\lambda(P) = \{\neg \mathrm{B}a \rightarrow_l b, \neg \mathrm{B}b \rightarrow_l a\}$. Remark that this is the theory T we studied in Examples 2 and 3. Hence we know that for each $x \in [0,1]$ there is a fuzzy stable expansion $\mathrm{Th}(S_x)$, with $S_x = \{I \mid I(a) \geq x \text{ and } I(b) \geq 1-x\}$. Hence for each $x \in [0,1]$ there is an answer set M_x such that $M_x(a) = \mathrm{Th}(S)(a) = \inf_{I \in S_x} I(a) = x$ and $M_x(b) = \mathrm{Th}(S)(b) = \inf_{I \in S_x} I(b) = 1-x$.

5 Related Work

Epistemic logic, the logic of epistemic notions such as knowledge and belief, is a major area of research in artifical intelligence. Von Wright's seminal work [22] is widely recognized as having initiated the formal study of epistemic logic as we know it today. Since then, various axiomatizations have been proposed, mainly in terms of possible-worlds semantics. An overview is given in [19]. Note that in general, epistemic logics may allow to model the beliefs of several agents, whereas autoepistemic logic is restricted to one's own beliefs. Autoepistemic logic has been important as an epistemic foundation for answer set set programming, which has also been studied from the angle of possibilistic logic [2], [4].

In recent years a variety of approaches to fuzzy answer set programming have been proposed, e.g. [3], [10], [21]. In [18] a fuzzy equilibrium logic was introduced, and a correspondence between fuzzy equilibrium logic models and answer sets of FASP programs was shown. Apart from this exception and our paper, it appears that little work has been done on nonmonotonic fuzzy logics nor about their relationship with fuzzy answer set programming.

We remark that fuzzy autoepistemic logic is also related to some work on fuzzy modal logics, see e.g. [9]. Another relevant paper is [1], where an epistemic modal logic is defined which is inspired by possibilistic logic. In this logic, interpretations are also sets of classical interpretations. Finally, there has also work been done on (finite) many-valued modal logics [11] and (finite) many-valued reflexive autoepistemic logic [12]. Instead of $[0,1]$, finite Heyting algebras are used for the space of truth values. Finitely-valued Gödel logic (truth values $\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots 1\right\}$) is a particular case of such algebras.

6 Conclusions

In this paper we have introduced a fuzzy version of autoepistemic logic, which can be used to reason about one's beliefs about the degrees to which properties are satisfied. We have shown that important properties of classical autoepistemic logic are preserved and that the relation between answer set programming and autoepistemic logic remains valid when generalizing to fuzzy logics. These results lead to a better comprehension of how to interpret fuzzy answer sets.

In future work, it would be interesting to see whether the implementation of classical autoepistemic logic by using quantified boolean formulas [5] can be extended to fuzzy logics using multi-level linear programming. If this is indeed the case, it could be used as a basis to implement fuzzy autoepistemic logic reasoners, as well as fuzzy answer set programming solvers. The general theory of fuzzy autoepistemic logic is also useful for abductive reasoning about theories with gradual propositions.

A Answer Set Programming (ASP)

We define a literal as either an atom or an atom preceded by not, the *negation-as-failure* operator. Intuitively, we say that not a is true if there is no proof to support atom a. If X is a set of atoms, we define not $(X) = \{\text{not } a \mid a \in X\}$. A *normal rule* is an expression of the form $a \leftarrow (\alpha \cup \text{not } (\beta))$, with a an atom and α and β sets of atoms. The atom a is called the *head* of the rule and $\alpha \cup \text{not } (\beta)$ (interpreted as conjunction) is the *body*.

A *normal program* P is a finite set of normal rules. The *Herbrand base* $\mathcal{B}_P$ of P is the set of atoms occuring in P. An *interpretation* I of P is any set of atoms $I \subseteq \mathcal{B}_P$. A *simple rule* is a normal rule without negation-as-failure. A *simple program* is a finite set of simple rules. If an interpretation I is the minimal model of P (i.e. the minimal interpretation such that $[r]_I = 1$ for each $r \in P$), then we say that I is the *answer set* of P. Thus, the answer set of a simple program P is the maximal set of atoms that can be deduced from P. For programs that are not simple, answer sets are defined using the Gelfond-Lifschitz reduct. Suppose P is a normal program, the *Gelfond-Lifschitz reduct* [7] of P w.r.t. the interpretation I is the set $P^I = \{a \leftarrow \alpha \mid (a \leftarrow (\alpha \cup \text{not } (\beta)) \in P, \beta \cap I = \emptyset\}$, which is a simple program. We then say that I is an *answer set* of P iff I is the answer set of P^I.

Note that simple programs have exactly one answer set, while normal programs can have 0, 1 or more answer sets.

References

1. Banerjee, M., and Dubois, D. A simple modal logic for reasoning about revealed beliefs. In *Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (2009), pp. 805–816.
2. Bauters, K., Schockaert, S., De Cock, M., and Vermeir, D. Possibilistic answer set programming revisited. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence* (2010).

3. Damásio, C. V., Medina, J., and Ojeda-Aciego, M. Sorted multi-adjoint logic programs: termination results and applications. In *Proceedings of the 9th European Conference on Logics in Artificial Intelligence* (2004), pp. 260–273.
4. Dubois, D., Prade, H., and Schockaert, S. Règles et méta-règles dans le cadre de la théorie des possibilités et de la logique possibiliste. In *Rencontres Francophones sur la Logique Floue et ses Applications* (2010), pp. 115–122.
5. Egly, U., Eiter, T., Tompits, H., and Woltran, S. Solving advanced reasoning tasks using quantified boolean formulas. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence* (2000), pp. 417–422.
6. Gelfond, M. On stratified autoepistemic theories. In *Proceedings of the Sixth National Conference on Artificial Intelligence* (1987), pp. 207–211.
7. Gelfond, M., and Lifschitz, V. The stable model semantics for logic programming. In *Proceedings of the Fifth International Conference and Symposiom on Logic Programming* (1988), pp. 1070–1080.
8. Hajek, P. *Metamathematics of Fuzzy Logic.* Springer, 2001.
9. Hajek, P. On fuzzy modal logics. *Fuzzy Sets and Systems 161*, 18 (2010).
10. Janssen, J., Schockaert, S., Vermeir, D., and De Cock, M. General fuzzy answer set programs. In *Proceedings of the International Workshop on Fuzzy Logic and Applications* (2009), pp. 353–359.
11. Koutras, C., Koletsos, G., and Zachos, S. Many-valued modal non-monotonic reasoning: Sequential stable sets and logics with linear truth spaces. *Fundamenta Informaticae 38*, 3 (1999), 281–324.
12. Koutras, C., and Zachos, S. Many-valued reflexive autoepistemic logic. *Logic Journal of the IGPL 8*, 1 (2000), 403–418.
13. Lifschitz, V., and Schwarz, G. Extended logic programs as autoepistemic theories. In *Proceedings of the Second International Workshop on Logic Programming and Nonmonotonic Reasoning* (1993), pp. 101–114.
14. Marek, W. Stable theories in autoepistemic logic. Unpublished note, Department of Computer Science, University of Kentucky, 1986.
15. Marek, W., and Truszczynski, M. Autoepistemic logic. *Journal of the Association for Computing Machinery 38*, 3 (1991), 587–618.
16. Moore, R. Semantical considerations on nonmonotonic logic. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence* (1983), pp. 272–279.
17. Moore, R. Possible-world semantics in autoepistemic logic. In *Proceedings of the Non-Monotonic Reasoning Workshop* (1984), pp. 344–354.
18. Schockaert, S., Janssen, J., Vermeir, D., and De Cock, M. Answer sets in a fuzzy equilibrium logic. In *Proceedings of the 3rd International Conference on Web Reasoning and Rule Systems* (2009), pp. 135–149.
19. Sim, K. Epistemic logic and logical omniscience: A survey. *International Journal of Intelligent Systems 12* (1997), 57–81.
20. Stalnaker, R. A note on non-monotonic modal logic. *Artificial Intelligence 64*, 2 (1993), 183–196.
21. Straccia, U. Annotated answer set programming. In *Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (2006).
22. von Wright, G. *An Essay in Modal Logic.* Studies in Logic and the Foundations of mathematics. Amsterdam: Nort-Holland Pub. Co., 1951.