Appeared in Akinori Abe and Yukio Ohsawa (Eds)
Readings in Chance Discovery
Advanced Knowledge International 2005

Cognitive Relevance and Chance Discovery

Xiaoping Chen

Multi-Agent Systems Lab, Department of Computer Science

University of Science and Technology of China

Hefei, 230026, China

xpchen@ustc.edu.cn

Abstract: We propose a formal theory to capture qualitatively the characteristics of utility-relevance and effect-uncertainty of chance. It sets a definite boundary between chance discovery and classical artificial intelligence planning: the former pursues partial realization of desire while the latter complete realization. The theory also provides a formal specification and deriving machinery for partial realization. The main formal properties and cognitive rationality of the theory are examined and justified. According to our theory, chance discovery is reduced to a class of computational problems, which can be solved on the basis of qualitative, e.g. common-sense, knowledge but cannot be solved by classical AI planning.

1 Introduction

As a new research area Chance Discovery has drawn increasing attention in recent years (Ohsawa, 2002; Abe, 2001; Prendinger and Ishizuka, 2001; Ji and Chen, 2002; Zhu and Chen, 2003). Some researchers focus on discovering chances by human beings, while others concentrate on automated chance discovery. There are two well accepted characteristics of chance in both contexts. (1) Utility-relevance. A chance is relevant to the interest of the agent who makes decision, either beneficial or deleterious. (2) Effect-uncertainty. The effect of a chance event/action is hard to be predicted or controlled, and thus it is difficult to be sure if the event/action is actually a chance. These two characteristics are essential to chance discovery, since they reflect the basic function and nature of chance so that any definition of chance must account for them. In this paper, we try to capture these characteristics and define, at least a very important sort of, chance discovery problems in a qualitative theory.

The first crucial issue to our task is to account for the effect-uncertainty. Usually chance is considered as something connected with randomness and measured in probability. This view makes chance discovery dependent on statistic data. However, a major assumption of chance discovery is that typically there are no such data available. In order to overcome this difficulty, we reduce the effect-uncertainty to a special kind of defective qualitative knowledge, i.e. those about only partial effect of action or event. This is equivalent to assume that an agent can only know or control partial effect of his/her action and the environment determines the other part of the whole effect. Thus an action can be regarded as a chance by an agent, if the agent only knows its partial effect. One can think the defective knowledge about partial effect of action reflects some kind of “qualitative randomness”. Consequently, an agent can rely on his/her qualitative knowledge, e.g. common-sense knowledge, in chance discovery.

Then consider the second crucial issue: how is the effect-uncertainty accounted for above connected to utility-relevance? First of all, we represent utility qualitatively as some logical formulas called desires (Cohen and Levesque, 1990; Chen and Liu, 1999). Each desire specifies a set of states that the agent wants to reach. Since an agent can only know or control partial effect of his/her action, generally he/she can only pursue partial satisfaction of a desire with this kind of actions. On the other hand, it seems that in everyday and social life human beings pursue partial satisfaction of his/her desires more often than not, since frequently their desires are *flexible*. When a human being states he/she desires something, usually he/she is willing to accept as alternatives some other things not logically equivalent to what he/she declares. Therefore, it is not only inevitable but also rational for an agent to content himself/herself with partial satisfactions of his/her desires by taking actions. Thus effect-uncertainty coincides with utility-relevance on the basis of partial satisfaction of desire through action.

The third basic issue is the difference between the chance discovery we account for above and the classical artificial intelligence (AI) planning, since the latter also assumes utility-relevance and effect-uncertainty. However, classical AI planning pursues complete realization of desire; when there are no actions to realize the desire completely, the planning problem is defined as no solutions (Ch.22, Nilsson, 1998). Therefore, we can distinguish chance discovery from classical AI planning according to their criteria of solutions: whether permitting partial satisfaction of desires. With this difference we set a definite boundary between these two classes of problems.

Now we can outline our theory as follows. The qualitative knowledge about partial effects of actions is captured by means of the "standard" definition of knowledge and action (Halpern and Moses, 1985; Parikh, 1980) with some simplification. In order to characterize and justify the partial realization of desires, we disassemble the desires of an agent into two sorts, explicit and implicit ones. The *explicit desires* are assumed given by the agent's declaration and closed under logical equivalence; the *implicit desires* are derived as desire-consequences of explicit desires, such that each of them corresponds to a partial realization of some explicit desire. A chance is defined as any action which the agent knows will realize some of his/her implicit desires, i.e. partially realize some of his/her explicit desires.

The major technical difficulty is the formalization of the desire-consequence or partial realization. We develop a new version of L_{m4c} as a formal specification and derivation mechanism for the desire-consequence, so that this consequence relation conforms to the rationale we set for it and its derivation is computationally tractable. The rationale for the desire-consequence is cognitive relevance, which distinguishes our theory from all of the previous ones.

In the next section, we present a framework for describing decision-making problems and distinguishing chance discovery from classical AI planning. We establish the new version of L_{m4c} and examine its main formal properties in section 3. In section 4 we discuss the rationality of the desire consequence in depth and make a definition of chance discovery problem. Finally in section 5 we draw some conclusions and point the way toward further development of the approach.

2 Cases Needing CD

First we specify a formal language to describe decision-making problems. Let L_B be a classical propositional language with the set of propositional symbols, *Atom* = {X_1, X_2 …}. Formulas of L_B are defined as usual: (1) any propositional symbol is a formula; (2) if φ and ψ are formulas of L_B, then $\neg\varphi$, $\varphi \vee \psi$, $\varphi \wedge \psi$, $\varphi \supset \psi$ and $\varphi \equiv \psi$ are formulas of L_B. Let L be an extension of L_B with a set of action term, *Action,* and modal operators $\boldsymbol{D_e}$ and $\boldsymbol{K}$. Any formula of L_B is also a formula of L. Besides, L includes any formula of the form $\boldsymbol{D_e}(\varphi)$ or $\boldsymbol{K}(\varphi / a)$, where φ is a formula of L_B and $a \in Action$. Let *Form* denote the set of formulas of L. Intuitively, $\boldsymbol{D_e}(\varphi)$ means that φ is one of the agent's explicit desires and $\boldsymbol{K}(\varphi / a)$ means that the agent knows φ will be true after he/she executes action a.

Usually, one needs a much more complicated language in the research of reasoning about action in order to treat the "compositions" of primitive actions and reason about their effects (Cohen and Levesque, 1990; Rao and Georgeff, 1991). But for the purpose of this paper, we needn't involve ourselves in the details and can simply employ the formulas of the form $\boldsymbol{K}(\varphi / a)$ to express the information in a rougher granularity. The exact meaning of any formula of L is given by *interpretations*.

Definition 1 (Interpretation**)** An interpretation is a structure $I = \langle S, V, D, K\rangle$, where S is a set of states, v is a mapping from *Atom* to the set of truth-value {1, 0}, $D \subseteq S$ ($D \neq \Phi$ & $D \neq S$) specifies the set of the agent's explicitly desired states, and K is a mapping from *Action* to 2^S, the power set of states.

Any formula φ of L being *true in a state s* under I, denoted by I, $s \models \varphi$, is defined as follows:

(1) $I, s \models X_i$ if $V(X_i) = 1$;

(2) $I, s \models \neg \varphi$ if $I, s \not\models \varphi$;

(3) $I, s \models \varphi \vee \psi$ if $I, s \models \varphi$ or $I, s \models \psi$;

(4) $I, s \models \varphi \wedge \psi$ if $I, s \models \varphi$ and $I, s \models \psi$;

(5) $I, s \models \varphi \supset \psi$ if $I, s \not\models \varphi$ or $I, s \models \psi$;

(6) $I, s \models \varphi \equiv \psi$ if $I, s \models \varphi \supset \psi$ and $I, s \models \psi \supset \varphi$;

(7) $I, s \models \boldsymbol{D_e}(\varphi)$ if $I, s \models \varphi \Leftrightarrow s \in D$;

(8) $I, s \models \boldsymbol{K}(\varphi / a)$ if $\forall s' \in K[a]$: $I, s' \models \varphi$.

I is a *model* of φ, denoted by $I \models \varphi$, if $I, s \models \varphi$ for all $s \in S$. φ is *valid under P,* a set of formulas, denoted by $P \models \varphi$, if any model of P is also a model of φ, where a model of P is a model of each $\psi \in P$. φ is *valid*, denoted by $\models \varphi$, if it is valid under the empty set; or equivalently, every interpretation is a model of φ.

The formulas of L_B are interpreted the same way as in the classical logic. $\boldsymbol{K}(\varphi / a)$ is interpreted in a way combining the "standard" theory of knowledge (Halpern and Moses, 1985) and PDL(Parikh, 1980) with some simplification. $K[a]$ specifies all of the states the agent knows he/she will arrive at after executing a. Usually $K[a]$ contains more than one state, since the agent does not possess complete knowledge about the environment or the effects of actions, and the agent only has *partial control* on the future---he/she cannot completely determine his/her future by taking an action. However, since φ is true in all of the states the agent thinks he/she will arrive at after executing a, his/her knowledge about φ's being true after executing a is certain. As mentioned in section 1, such knowledge can only specifies partial effects of actions. The explicit desire is interpreted as a non-normal modal operator, similar to that in (Konolige and

Pollack, 1993). We will give a detailed explanation for this choice later in section 4. This interpretation is a special case of the minimal model semantics (Chellas, 1980). But we need a constraint: $D \neq \Phi$ & $D \neq S$; this refuses any contradictions or tautologies to be explicit desires. By the definition, given $\boldsymbol{D_e}(\varphi)$, for any ψ $\boldsymbol{D_e}(\psi)$ holds if and only if ψ is logically equivalent to φ. This means that φ is the unique explicit desire of the agent and thus the sufficient and necessary condition for what the agent explicitly desires. This is only an assumption for simplicity. The main results of this paper are valid in general case where the agent has a finite number of explicit desires.

Now we can give a definition of decision-making problem from the point of view of *reasoning about action* in the framework given above.

Definition 2 (Decision-making problem**)** $P = \langle \boldsymbol{D_e}(\varphi), \{\boldsymbol{K}(\varphi_a / a) | a \in Action\}\rangle$ is called a *decision-making problem*, if $\boldsymbol{D_e}(\varphi)$ is the explicit desire and $\{\boldsymbol{K}(\varphi_a / a) | a \in Action\}$ the knowledge of the agent.

What is a solution to any given decision-making problem? In AI literature, the problem defined above is regarded as a (classical) planning problem and any action a such that $\models \boldsymbol{D_e}(\varphi) \supset \boldsymbol{K}(\varphi / a)$ is defined as a solution to the problem (Ch.22, Nilsson, 1998), called hereafter a CAIP solution for short. However, let's consider the relationships between the agent's explicit desire and his/her knowledge about the effect of any action. Given a decision-making problem P, an interpretation I and an action a. There are four possible cases, as shown in Figure1.

Case 1: $K[a] \subseteq$ D. In this case each state the agent knows he/she will possibly arrive at by taking action a is an explicitly desired state. It follows from the definitions above that the agent knows his/her explicit desire will be satisfied after executing action *a* under *I*; i.e., $I \models \boldsymbol{D_e}(\varphi) \supset \boldsymbol{K}(\varphi / a)$. We call such an action *a sufficient for P*.

Case 2: $D \subseteq K[a]$ & $D \neq K[a]$. In this case some states the agent knows he/she possibly will arrive at by taking action *a* are not his/her explicitly desired states. We call such an action *a unsafe* to *P*. The agent does not know that his/her explicit desire can be realized by taking action *a*, or equivalently, $I \models D_e(\varphi) \supset \neg \boldsymbol{K}(\varphi / a)$. Note that in this case the agent has no definite knowledge about the effect of action *a*.

Case 3: $K[a] \cap D \neq \Phi$ & $K[a] \cap (S \setminus D) \neq \Phi$.The case is similar to case 2 except, even worse, that some of the desired states cannot be realized by taking action *a* according to the agent's knowledge. Such an action *a* is called *unreliable* to *P*. Neither in this case has the agent any definite knowledge about the effect of action *a*.

Case 4: $K[a] \cap D = \Phi$, or equivalently, $I \models D_e(\varphi) \supset \boldsymbol{K}(\neg \varphi / a)$. The agent in this case knows that it is impossible to realize his/her explicit desire by executing the action *a*. It is different from case 2 and 3 that the agent in this case has definite knowledge about the fact that the action *a* is not a solution to *P*. We call *a* a *conflictive* action to *P*.

According to the criterion of classical AI planning, an action *a* is a solution to the problem *P*, if and only if in all interpretations *a* is sufficient for the explicit desire. This implies that if all of the actions are unsafe, unreliable or conflictive to the problem, there are no CAIP solutions. It is well known in the area of AI planning that generally there is no algorithm that can give any definite answer to a decision-making problem if there is no CAIP solution to the problem; more accurately, the decision-making problem under the criterion of classical AI planning is semi-decidable. On the other hand, whether there is a CAIP solution to a decision-making problem depends on whether the agent is omniscient (e.g. the agent's knowledge and ability of acting are complete). If the agent is not omniscient, then there are no CAIP solutions to the problem. Unfortunately, it is the case in most real-world problems. This implies that a classical AI planning algorithm, which tries to find a CAIP solution, can do nothing for most real-world decision-making problems. In this sense an appropriate definition of classical AI planning problem should not cover those cases where there is no CAIP solution. In fact, most research on AI planning concentrates on finding CAIP solutions effectively. This observation leads to following definition.

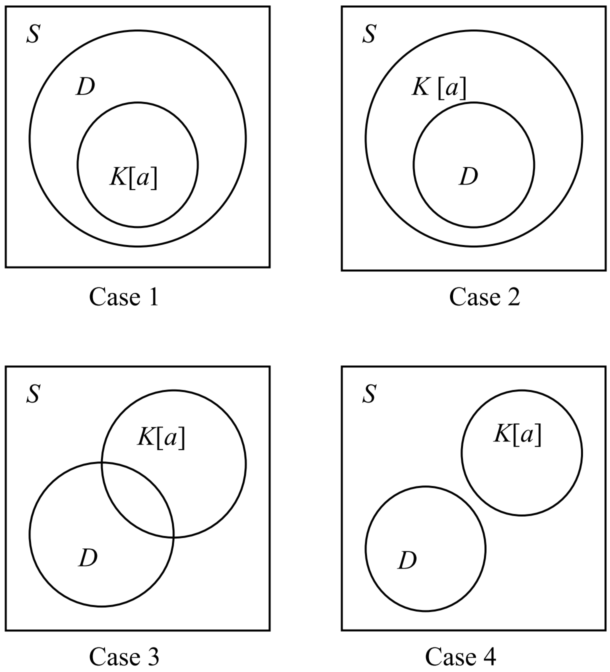

Figure 1

Definition 3 (Planning problem**)** Any decision-making problem $P = \langle \boldsymbol{D_e}(\varphi), \{\boldsymbol{K}(\varphi_a / a) \mid a \in Action\}\rangle$ is a *planning problem*, if there is a *CAIP solution* to *P*, i.e. there is some $a \in Action$ such that $P \models \boldsymbol{D_e}(\varphi) \supset \boldsymbol{K}(\varphi / a)$.

Much more importantly, what is a solution to a decision-making problem if there is no CAIP solution? Hereafter we call such problems *hard decision-making problems.* It is easy to see that there is no answer to this question if one insists on CAIP solutions to hard decision-making problems. Hence we have to give up CAIP solutions and establish a new criterion for solving hard decision-making problems.

We claim that chance discover provides a new approach to solving, at least some kind of, hard decision-making problems. But first we must give a new criterion of solutions. As the basis of such a criterion, we provide a formal specification for desire consequences which reflects the flexibility of desire.

3 Desire Consequence

As mentioned in section 1, we assume that the agent's desire is flexible, i.e. he/she has some implicit desires, each of which is implied by some but not logically equivalent to any explicit desire of the agent. In fact, it is neither cognitive rational nor computational feasible to represent flexible desires completely explicit, e.g. declaring them one by one. Hence we disassemble flexible desires of an agent into two sorts, explicit and implicit ones, and define a desire consequence relation to derive implicit desires from the explicit ones. With the formal specification of this desire consequence we can also study and clarify the cognitive and formal characteristics of implicit desires. We introduce the specification and examine its formal properties in this section, and then explain the intuition behind the definition in the next section.

Let L_E, the language of L_{m4c}, be an extension of L with additional operators $\rightarrow$ and $\boldsymbol{D_i}$ representing desire consequence and implicit desire, respectively. L_E contains all of the formulas of L and the formulas in the form $\varphi \rightarrow \psi$ or $\boldsymbol{D_i}(\varphi)$ where $\varphi, \psi \in L_B$. In the original version of L_{m4c}, its semantic interpretation is given over set T = {t, f, 1, 0} (Chen and Liu, 1999). For the purpose of this paper, however, we need to develop a new semantic interpretation.

In order to simplify the descriptions in this section, we introduce a slightly different notation. A *literal* is a propositional symbol or the negation of a propositional symbol. A *state* is a consistent set of literals such that x or $\neg x$ is in it for any $x \in$ Atom. Given an interpretation I. The set of states in I is still denoted by S. For any $\varphi \in L_B$ and $s \in S$, let $s[\varphi] = 1$ if $I, s \models \varphi$; $s[\varphi] = 0$, otherwise. Let $|\varphi| = \{s \mid s[\varphi] = 1\}$. Any $s \in |\varphi|$ is called a φ-*state*. It follows that $s[D_e(\varphi)] = 1$ if $D_e = |\varphi|$ and $s[D_e(\varphi)] = 0$ otherwise for all $s \in S$. A *partial state* is a consistent set of literals. Let S_P denote the set of partial states. Clearly, we have $S \subseteq S_P$. For any $s \in S$ and $r \in S_P$, if $r \subseteq s$ then r is called a *partial state of s*, denoted by $r \in S_P(s)$, and $s - r$ (the set that contains all literals belonging to s but not r) is called the *context of r in s*. If there is some s such that c is the context of r in s, then c is called a *context* of r. Any r' is called a *variant* of r, if there is some context c of r such that $r' \cup c \in S$. Given any $r \in S_P(s)$ and a desire, explicit or implicit, denoted by $\boldsymbol{D}(\varphi)$. r is called *cognitively related to* $\boldsymbol{D}(\varphi)$ in s, denoted by $r \in R(\varphi \mid s)$, if there is a variant r' of r such that $(r' \cup s{-}r)[\varphi] \neq s[\varphi]$, i.e. the change of the truth value of r can cause the change of the truth value of φ. For example, we have $\{X_1\} \in R(X_1 \wedge X_2 | \{X_1, X_2, X_3, \ldots\})$ and $\{X_3\} \notin R(X_1 \wedge X_2 | \{X_1, X_2, X_3, \ldots\})$. Intuitively, $r \in R(\varphi \mid s)$ means that the truth value of φ in s depends on that of the part r of s. Any r is called *cognitively related to* $\boldsymbol{D}(\varphi)$, denoted by $r \in R(\varphi)$, if there is some state s such that $r \in R(\varphi \mid s)$.

Given any $s \in S$ and $r \in R(\varphi \mid s)$. If there is no $r' \subseteq r$ such that $r' \neq r$ and $r' \in R(\varphi \mid s)$, then r' is called a *minimal cognitively-related part of r* and denoted by $r \in M(\varphi \mid s)$. Any φ is non-trivial, if it is not a contradiction or a tautology. For any $\boldsymbol{D}(\varphi)$ where φ is non-trivial, if there is some $s \in |\varphi|$ such that $r \in$

M(φ | s), then r is called *strong cognitively-related* to ***D***(φ), denoted by $r \in$ M(φ). Obviously, a cognitively related partial state r of a state s may contain a part not cognitively related to s. For instance, $\{X_1, X_3\} \in R(X_1 \wedge X_2)$ and $\{X_3\} \notin R(X_1 \wedge X_2)$. However, it is not the case for minimal cognitively-related part. Hence any strong cognitive-related partial state is a substantial part of any φ-state if the agent's desire is ***D***(φ). An agent with desire ***D***(φ) commits and only commits himself/herself to all of the partial states which are strongly cognitively-related to ***D***(φ). Based on this observation, we define desire consequence as follows.

Definition 4 (Desire consequence) For φ, $\psi \in L_B$, ψ is a *desire consequence* of φ, denoted by $|= \varphi \to \psi$, if and only if ψ is non-trivial and for any $s \in |\psi|$ there is an $s' \in |\varphi|$ such that M(ψ | s) $\subseteq$ M(φ | s').

Let $|= \varphi \leftrightarrow \psi$ denote $|= \varphi \to \psi$ and $|= \psi \to \varphi$. Some basic instances and counter-instances of the desire consequences of L_{m4c} are as follows.

Property 5

(P1-1) $\models X_1 \wedge X_2 \to X_1$

(P1-2) $|\neq (\neg X_1 \wedge X_1) \to X_1$

(P1-3) $|= X_1 \wedge (X_1 \vee X_2) \to X_1$

(P1-4) $|\neq X_1 \wedge (X_1 \vee X_2) \to (X_1 \vee X_2)$

(P1-5) $\models X_1 \wedge (X_2 \vee X_3) \to X_1 \vee (X_2 \wedge X_3)$

(P2-1) $|= (X_1 \vee X_2) \to X_1$

(P2-2) $|\neq (\neg X_1 \vee X_1) \to X_1$

(P3-1) $\|\neq X_1 \to (X_1 \vee X_2)$

(P4-1) $\models (X_1 \wedge X_2) \leftrightarrow (X_1 \vee X_2)$

(P5-1) $\models \varphi \wedge (\psi \wedge x) \leftrightarrow (\varphi \wedge \psi) \wedge x$

(P5-2) $\models \varphi \vee (\psi \vee x) \leftrightarrow (\varphi \vee \psi) \vee x$

(P6-1) $\models \varphi \wedge \psi \leftrightarrow \psi \wedge \varphi$

(P6-2) $\models \varphi \vee \psi \leftrightarrow \psi \vee \varphi$

(P7-1) $\models \varphi \wedge (\psi \vee x) \leftrightarrow (\varphi \wedge \psi) \vee (\varphi \wedge x)$

(P7-2) $\models \varphi \vee (\psi \wedge x) \leftrightarrow (\varphi \vee \psi) \wedge (\varphi \vee x)$

(P8-1) $\models \neg(\varphi \wedge \psi) \leftrightarrow (\neg \varphi \vee \neg \psi)$

(P8-2) $\models \neg(\varphi \vee \psi) \leftrightarrow (\neg \varphi \wedge \neg \psi)$

(P9-1) $\models \neg\neg \varphi \leftrightarrow \varphi$

(P10-1) $\models \varphi \leftrightarrow \varphi \wedge \varphi$

(P10-2) $\models \varphi \leftrightarrow \varphi \vee \varphi$.

Obviously, the desire consequence defined above is quite different from previous ones. Its most outstanding feature is free from side effects, as (P3-1) shows. A second feature is that this consequence is not a restriction of any classical logical consequence, since (P2-1) and (P4-1) are not valid in any previous one. With these distinguishing features, however, the consequence relation still conforms to the principle of logical equivalence as indicated by the instances from (P5-1) to (P10-2).

Theorem 6 (Reflexivity and Transitivity) For any non-trivial φ, x, $\psi \in L_B$ we have

(1) $\models \varphi \rightarrow \varphi$;

(2) if $\models \varphi \rightarrow x$ and $\models x \rightarrow \psi$, then $\models \varphi \rightarrow \psi$.

This theorem shows that the desire consequence is an inferential one and the following theorems show it is computable and its decision is also tractable.

Theorem 7 (Decidability) The desire consequence is decidable, i.e., there exists an algorithm such that for any non-trivial $\varphi, \psi \in L_B$, it returns "yes" if $\models \varphi \to \psi$ and "no" if $\not\models \varphi \to \psi$.

Theorem 8 (Finiteness) For any non-trivial φ, it has only a finite number of substantially different desire consequences.

Now we give the definition of implicit desires.

Definition 9 (Implicit desire) Any ψ is an *implicit desire* with respect to $\boldsymbol{D_e}(\varphi)$ in P, denoted by $P \models \boldsymbol{D_i}(\psi \mid \varphi)$, if $\models \varphi \to \psi$.

Therefore, for any explicit desire $\boldsymbol{D_e}(\varphi)$ there are a finite number of implicit desires, each of them can be derived from the explicit desire by an algorithm in a finite number of steps. In fact, there is an algorithm generating all of the implicit desires from any given explicit desire in a finite number of steps.

4 Partial realization

In most BDI theories (explicit) desire is interpreted in the same way as knowledge. It is well known, however, that this interpretation causes so called "side-effect problem" (Bratman, 1987; Linder *at al*, 1995). For example, it follows in this semantics that

(SE) $\models \boldsymbol{D_e}(\varphi) \supset \boldsymbol{D_e}(\varphi \vee \psi)$

for any φ and ψ. This is counterintuitive and harmful to many applications. Suppose an agent wants to be wise. Then it follows according to the semantics that the agent wants to be wise or stupid. As a partial solution to this problem we interpret the explicit desire as a non-normal modal operator. This definition rejects side effects such as (SE). However, it is inadequate to define all desires as such, since that would imply inflexibility of desire: an agent desires anything if and only if it is logically equivalent to something the agent declares. Therefore, we introduce desire consequence and derive implicit desires with it from explicit ones. A remarkable feature of our treatment is that the implicit desires are free from side-effects.

Consider the instance of (SE) where $\varphi = X_1$ and $\psi = X_2$. Clearly $s = \{X_1, X_2, X_3, \ldots\}$ is both a φ-state and a $\varphi \vee \psi$-state. However, the partial state $r = \{X_2\}$ of s is not cognitively related to φ, although it is to the classical desire consequence $\varphi \vee \psi$. That is to say, (SE) does not preserve the cognitive relevance in the sense that it can introduce cognitively irrelevant parts into the desire consequences. For example, “being stupid” is not cognitively related to “being wise”, but it is to “being wise or stupid”. Thus the classical desire consequence “being wise or stupid” of “being wise” contains and introduces the cognitive irrelevant part “being stupid”. This is the major cause of side-effect problem. Therefore desire consequences should preserve cognitive relevance in order to get rid of side-effects.

The most important property of the desire consequence we defined is that the implicit desires preserve the cognitive relevance with respect to the explicit desires. This property gives a strong justification to our theory.

Corollary 10 (Cognitive relevance) For any P and $\varphi, \psi \in L_B$, if $P \models \boldsymbol{D_e}(\varphi)$ and $P \models \boldsymbol{D_i}(\psi \mid \varphi)$, then $R(\psi) \subseteq R(\varphi)$.

The intuition behind this principle and the rationale of desire consequence can be explained as follows. Suppose $\boldsymbol{D}(\varphi)$ is a desire of the agent. This means that the agent wants to reach any φ-state $s \in |\varphi|$. All of these φ-states have no difference for this purpose since each of them satisfies the agent's desire. However, for any φ-state s, only some parts of s are cognitively related to the agent's desire. These are exactly those in $\mathrm{M}(\varphi | s)$. Any $s \in |\varphi|$ can be considered as an *instance realization* of φ and any $r \in \mathrm{M}(\varphi | s)$ a *substantial part* of this instance realization.

Now suppose $\boldsymbol{D_e}(\varphi)$ and $\models \varphi \rightarrow \psi$. Then according to the definition, for any $s' \in |\psi|$ there is an $s \in |\varphi|$ such that $\mathrm{M}(\psi | s') \subseteq \mathrm{M}(\varphi | s)$. In other words, any substantial part of an instance realization of ψ is also a substantial part of some instance realization of φ. This means that any instance realization of ψ is also a *partial instance realization* of φ with respect to the agent's desire. In other words, if the agent desires to reach φ, then he/she must desire to reach ψ as well. In this sense ψ is cognitively implied by $\boldsymbol{D_e}(\varphi)$, i.e. an implicit desire *wrt* the latter.

Introducing implicit desires opens a new possibility to solving hard decision-making problems. Suppose there is no CAIP solution to a decision-making problem, i.e. there is no action realizing the agent's explicit desire. If only the agent's desire is flexible, he/she can try to realize some of implicit desires instead of explicit ones. This provides a new criterion of decision-making, by which the agent gains chances for solving hard decision-making problems. Hence if a hard decision-making problem can be solved under this criterion, we consider it to be a chance discovery problem.

Definition 11 (Chance discovery problem) Any hard decision-making problem $P = \langle\ \boldsymbol{D_e}(\varphi),\ \{\boldsymbol{K}(\varphi_a\ /\ a) \mid a \in Action\}\rangle$ is a *chance discovery problem*, if there is some $\psi \in L_B$ and $a \in Action$ such that $P \models \boldsymbol{D_i}(\psi \mid \varphi) \supset \boldsymbol{K}(\psi / a)$.

Note that this definition embodies both utility-relevance and effect-uncertainty. Also note that there are many problems which are solvable in chance discovery but not in planning. For example, suppose an agent wants to win both a champion and a runner-up in some competition, i.e. the agent holds $\boldsymbol{D_e}(X_1 \wedge X_2)$. But he knows that he can only win a champion, i.e. there is some $a \in Action$ such that $\boldsymbol{K}(X_1 / a)$ belongs to his knowledge base but there is no a such that $\boldsymbol{K}(X_2 / a)$ belongs to his knowledge base. Then the agent cannot realize his explicit desire; in other words, there are no solutions to this planning problem. However, the agent can take the action a to realize his implicit desire $\boldsymbol{D_i}(X_1 \mid X_1 \wedge X_2)$ and thus finds a solution to this hard decision-making problem. In this sense the agent gains an opportunity by changing the criterion or the type of decision-making problems.

There is a second sense in which the agent gains an opportunity in the way described above. Suppose that the agent does not know there is an action realizing his second sub-goal, winning the runner-up, before his taking the action a to realize his first sub-goal, winning the champion. But after he executes a, he acquires more information and knows his second sub-goal is also reachable. Thus the agent actually realizes both sub-goals of his desire separately with the help of sequent knowledge.

However, it is easy to see that an agent can possibly come up against some risks this way. One can explain the reasons with the help of figure 1. But this is inevitable cost for chance discovery and coincides with the experience of everyday life. Our theory can also be employed to distinguish opportunities with risks.

Definition 12 (Opportunity/risk) Given any chance discovery problem $P = \langle \boldsymbol{D_e}(\varphi), \{\boldsymbol{K}(\varphi_a / a) | a \in Action\}\rangle$. An action $a \in Action$ is an opportunity to P, if there is some $\psi \in L_B$ such that $\models \boldsymbol{D_i}(\psi | \varphi) \supset \boldsymbol{K}(\psi / a)$. An action $a \in Action$ is a risk to P, if there are some $\psi \in L_B$ such that $\models \boldsymbol{D_i}(\psi | \varphi) \supset \boldsymbol{K}(\neg \psi / a)$.

Clearly, an action can be possibly both an opportunity and a risk to a chance discovery problem at the same time. In the example above, any action causing the agent to win the champion and lost the runner-up has such a twofold function. This reflects an aspect of effect-uncertainty of chance.

All of the opportunity actions can be divided into three mutually disjoint classes, consisting of unsafe, unreliable and conflictive actions to the corresponding classical AI planning problem, respectively. Maybe to one's surprise, a conflictive action can be a chance. But it is the case in some hard decision-making problems. For example, suppose $\boldsymbol{D_e}(p \wedge q)$ and $\boldsymbol{K}(p \wedge \neg q / a)$ hold in some hard decision-making problem P. Then the agent can choose implicit desire $\boldsymbol{D_i}(p | p \wedge q)$ as an alternative to his/her explicit desire and thus a becomes a chance, since $\models \boldsymbol{D_i}(p | p \wedge q) \supset \boldsymbol{K}(p / a)$. Of course, this action a is also a risk to the problem since $\models \boldsymbol{D_i}(q | p \wedge q) \supset \boldsymbol{K}(\neg q / a)$. In fact, in all of the cases the alternative implicit desires are formed by weakening the explicit desire so that the agent knows they can be realized by taking some actions.

5 Conclusion

In a decision-making problem an agent tries to find an action to satisfy his/her desire. However, there can be different criteria for "satisfying a desire". In a classical AI planning problem, the standard criterion is to reach a state in which the agent's desire is satisfied completely. Our analysis in section 2 reveals that most real-world decision-making problems cannot be solved according to this criterion. This indicates the necessity of introducing chance discovery into real-world decision-making problems. Based on the assumption of flexible desire and the separation of implicit desires from explicit ones, we establish a new criterion of decision-making and define chance discovery problems as those hard decision-making problems that can be solved according to the new criterion. In this way we set a definite boundary between classical AI planning and chance discovery for the first time.

According to our division, there are some significant and remarkable differences between these two classes of problems. Firstly, in classical AI planning an agent pursues "complete logical solutions" in the sense that any solution action will bring the agent to a state where the agent's explicit desire is completely satisfied, while in chance discovery an agent pursues partial satisfaction of his/her explicit desire. Secondly, the fundamental assumption of classical AI planning is that the agent's desire is inflexible, while that of chance discovery flexible. This reflects our experience of everyday life---flexibility helps one comes across more chances, though maybe more troubles as well. Thirdly, an agent in planning only deliberates within the range of knowledge, while in chance discovery the deliberation involves both knowledge and desires. In this sense the chance discovery problems defined in this paper is more complex than classical planning problems.

Technically, the key concepts of our theory and the new criterion of decision-making are captured and characterized in a new version of L_{m4c}, which provides a formal specification and decidable deriving mechanism for flexible desire consequences. This specification constitutes a formal foundation for qualitative discovery of chance, just like classical logic for classical AI planning. We show that this system has satisfactory properties and a cognitively rational basis. Most remarkably, the desire consequence relation conforms to the principle of cognitive relevance or desire-preserving, while previous theories do not provide specifications of desire consequence which conform to the same principle and thus suffer from side-effect problems. It is also deserved to point out that there is a fast algorithm for the derivation of L_{m4c} (Zhou, *et al*, 2001). As a result, chance discovery can be regarded as searching for an action realizing implicit desire, just like planning as searching for an action realizing explicit desire. This means that chance discovery becomes a class of pure computational problems once our definition is accepted.

Although we try to establish a more solid foundation to the previous endeavor in the same approach, there is a lot of work to do in the future. Since the version of L_{m4c} given in the paper confines to propositional language and many decision-making problems can only be represented in first-order language, we need develop its first-order version. Though partial effects of actions can be represented in standard language of knowledge and action, this representation do not support thorough analysis on effect-uncertainty. Thus we need to establish more powerful frameworks and go further into the depth of chance discovery.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (60275024).

References

Akinori, A. (2001), "The role of abduction in Chance Discovery", *Proceedings of SCI-2001*, Vol. VIII, pp. 400-405.

Bratman, M.E. (1987), *Intentions, Plans and Practical Reason*, Harvard University Press: Cambridge, MA.

Chellas, B.F. (1980), *Modal Logic: An Introduction*, Cambridge University Press.

Chen, X. and Liu, G. (1999), “A logic of intention”, *Proceedings of IJCAI-99*, Morgan Kaufmann, pp. 172-177.

Cohen, P.R. and Levesque, H.J. (1990), “Intention is choice with commitment”, *Artificial intelligence* ***42***, pp. 213-261.

Ji, C. and Chen, X. (2002), “Chance Discovery in a BDI perspective of planning: anticipation, participation and correlation”, *Working Notes of the 2nd International Workshop on Chance Discovery*, Tokyo.

Halpern, J.Y. and Moses, Y.O. (1985), “A guide to the modal logics of knowledge and belief”, *Proceedings of IJCAI-85*, Los Angeles, CA.

Konolige, K. and Pollack, M.E. (1993), “A representationalist theory of Intention”, *Proc IJCAI-93*, 1993, pp. 390-395.

Linder, B van. Hoek, W van der. and Meyer, J.-J. Ch. (1995), “Formalizing motivational attitudes of agents”, In J. P. Muller and M. Tambe (eds.), *Proc. of IJCAI’95 Workshop*(ATAC), 1995, pp. 17-32.

Nilsson, N.J. (1998), *Artificial Intelligence—A New Synthesis*, Morgan Kaufmann.

Ohsawa, Y. (2002), “Chance discoveries for making decisions in complex real world”, *New Generation Computing* (Springer Verlag and Ohmsha), Vol. 20, No. 2, pp. 143-163.

Parikh, R. (1980), “Propositional logics of programs: systems, models, and complexity”, *Proceedings of the 7th Symposium on Principle of Programming Languages*, Las Vegas.

Prendinger, H. and Ishizuka, M. (2001), “Some methodological considerations on Chance Discovery”, *Proc. of SCI-2001*, Vol. VIII, pp. 425-428.

Rao, A.S. and Georgeff, M.P. (1991), “Modeling rational agents within a BDI-architecture”, *Proc of KR-91*, pp. 473-484.

Zhou, J., Zhang, R. and Chen, X. (2001), “A decision algorithm for L_{m4c} consequences”, *Mino-Micro Computer Systems*. Vol.22, No.10.

Zhu, S. and Chen, X. (2003), “Relevance in Chance Discovery”, *Working Notes of HCI- 2003 Workshop on CD&CM.* Greece.