OCC Theoryに基づくエージェントの感情表現の論理モデルについて

奈良女子大学 人間文化研究科 情報科学専攻
池之内彰子

2014年5月27日

研究の背景

- 人との対話を実現させる上で、より人間に近いエージェントが求められている
- エージェントへの感情の付与が重視されている
- 感情が振る舞いに与える影響を考慮する必要がある

研究の背景

- 人との対話を実現させる上で、より人間に近いエージェントが求められている
- エージェントへの感情の付与が重視されている
- 感情が振る舞いに与える影響を考慮する必要がある

⬇

感情を生起し、自力で行動を選択できる
BDIエージェントの実現

BDIモデル

- Belief(**信念**):**情報**
- Desire(**願望**):**願望**
- Intention(**意図**):**目標**

以上3種類の心的パラメータを明示的に持ち、
これらの心的パラメータを保持・更新することで
意思決定を行い、目標を達成するように振る舞う
エージェントのモデル。

OCC Theory

- Ortony, Clore, Collinsが提唱(1988年)
- 心理学的見地を基にした22種類の感情タイプをモデル化した理論
- 人間の包括的な感情を形式化
- 感情の特徴付けが明確

BDIモデルと OCC Theoryの親和性

- BDIモデル
  - 信念・願望・意図の3つの心的パラメータを用いた人間の合理的行動をモデル化したもので、BDI logicという**論理モデル**を持つ
- OCC Theory
  - 信念や願望などの心的状態を用いた感情の特徴付けが明確で、その感情の特徴付けを**論理モデル**で表現可能な形で行える

どちらも**論理モデル**を利用可能であるため、**親和性が高い**

BDIモデルとOCC Theoryを用いてエージェントの感情を扱っている研究

- OCC Theoryにおける感情の形式化
  「A logical formalization
  of the OCC theory of emotions」(2009, Adam他)
  - BDIモデルの形式的記述のための論理体系であるBDI logicに対し、新たなオペレータを導入して拡張し、OCC Theoryで扱われる感情の定義を論理式として形式化することで、BDIモデルに取り込んだ。

現在までの成果

OCC Theoryで定義されている22種類の感情を、Adamらの形式化を踏まえJasonで実装

- エージェントシステムに感情表現を組み込める汎用的な手法として利用可能に(2012, 清水・2013, 池之内)
- 汎用的に利用可能なライブラリとして提供(2014, 山根)

※Jasonとは、BDIアーキテクチャを実装した、エージェントの記述言語の処理系

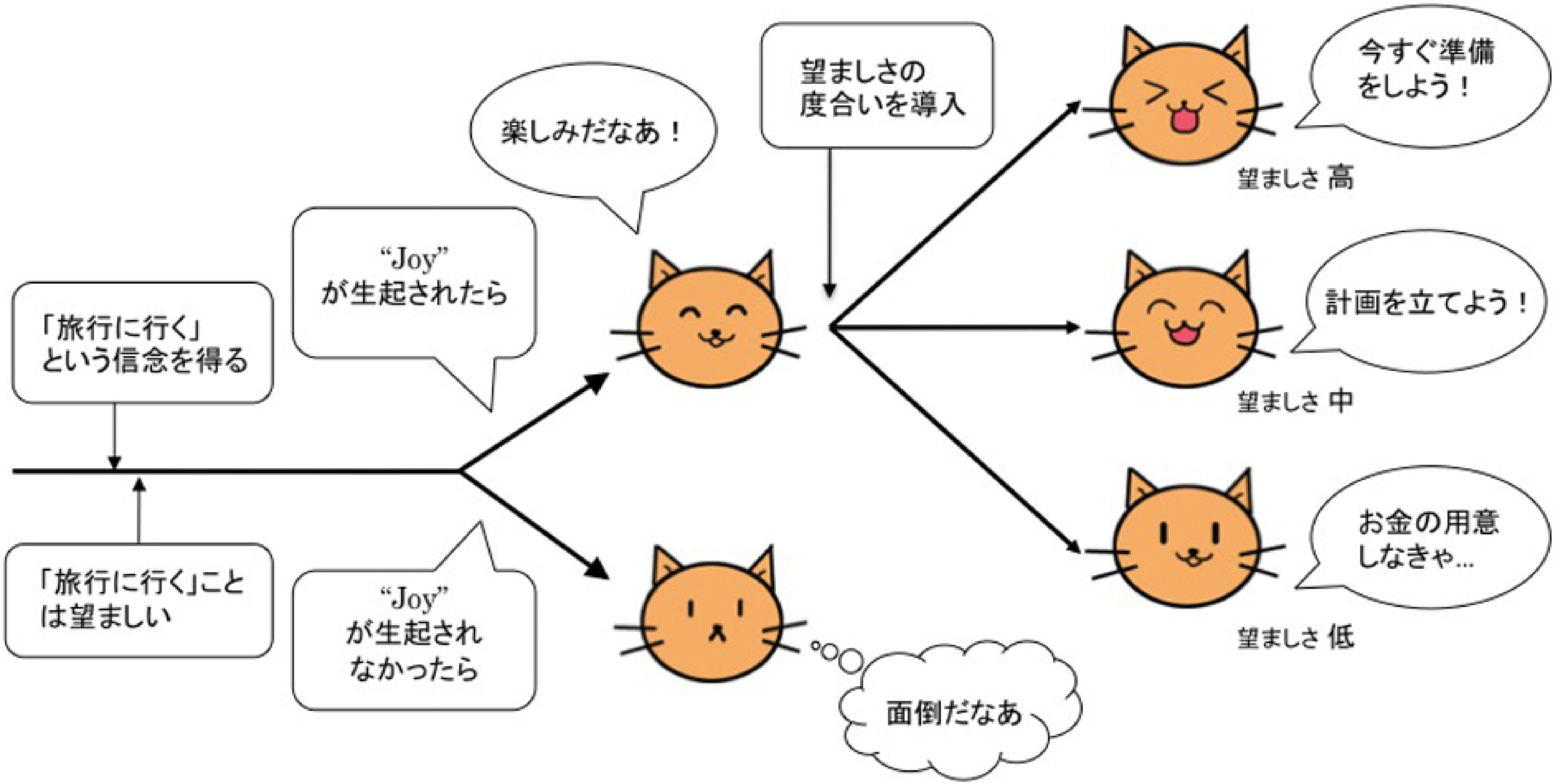
課題

- 感情の度合い(強さ)について考慮していない
- 生起した感情は時間が経過しても消滅しない

課題

- 感情の度合い(強さ)について考慮していない

Adam らの感情の形式化に、OCC Theory で定義されている “感情の強さに影響を与える変数” を導入し、
度合い付きの感情を表現できる論理体系を考案


「旅行に行く」
という信念を得る
「旅行に行く」こと
は望ましい
“Joy”
が生起されたら
“Joy”
が生起され
なかったら
楽しみだなあ！
面倒だなあ
望ましさの
度合いを導入
今すぐ準備
をしよう！
望ましさ 高
計画を立てよう！
望ましさ 中
お金の用意
しなきゃ...
望ましさ 低

構文論

- 命題記号の有限集合をSとする。命題記号は論理式
- エージェントを表す記号の有限集合$\mathcal{A}$
- ϕ, ψが論理式ならば、$\neg\phi$, $\phi \vee \psi$も論理式
  また、$\wedge$, $\supset$, $\leftrightarrow$の省略形を以下のように定義する。
  - $\phi \wedge \psi := \neg(\neg\phi \vee \neg\psi)$
  - $\phi \supset \psi := \neg\phi \vee \psi$
  - $\phi \leftrightarrow \psi := (\phi \supset \psi) \wedge (\psi \supset \phi)$

- ϕが論理式、$a \in \mathcal{A}$、$d(desirability)$ が$0 \leq d \leq 1$を満たす実数のとき、$Des^a_{=d}\phi$は論理式であるとする
  - $=$の他に$\geq$, $>$, $\leq$, $<$, $\neq$なども考えられる
    例えば$Des^a_{\geq 0.6}\phi$は「エージェントaにとってイベントϕの望ましさは0.6以上」を表すことになる

- ϕが論理式、$a \in \mathcal{A}$、$\ell(likelihood)$が$0 \leq \ell \leq 1$を満たす実数のとき、$Prob_\ell^a \phi$は論理式であるとする
  - 省略形として$Prob_{1.0}^a \phi := Bel^a \phi$ と定義する
    直感的な意味は、それぞれ
    - $Prob_\ell^a \phi$が「エージェントaは、見込みℓで、イベントϕが起こりそうだと信じている」
    - $Bel^a \phi$が「エージェントaは、イベントϕが起こると信じている」

    となる

- ϕが論理式ならば、$G\phi$も論理式
  - 省略形として$\neg G \neg \phi := F\phi$と定義する
    直感的な意味は、それぞれ
    - $G\phi$が「イベントϕはこれから真になる」
    - $F\phi$が「イベントϕは真である、もしくはこれから近い未来に真になる」

    となる

- ϕが論理式ならば、$H\phi$も論理式
  - 省略形として$\neg H \neg \phi := P\phi$と定義する

    直感的な意味は、それぞれ
    - $H\phi$が「イベントϕは過去常に真だった」
    - $P\phi$が「イベントϕは真である、もしくは真だった」

    となる
- $P\phi$に、努力の程度を表す変数$e(effort)$を付与する

  ϕが論理式、$a \in \mathcal{A}$、eが$0 \leq e \leq 1$を満たす実数のとき、$P^a_e\phi$は論理式であるとする

構造

- 可能世界の集合 $W (\neq \emptyset)$
- 各世界での命題記号に対する真偽の割り当て

$$V : W \times S \to \mathbb{B}$$

- $\mathcal{A}$の各要素aに対し、$W \times W$から$[0,1]$への関数

$$R_B^a : W \times W \to [0,1]$$

ただし任意の$a \in A,\ w \in W$に対し、

$$\sum_{w' \in W} R_B^a(w, w') = 1$$

であること。また、R_B^aは、推移的かつSerialかつEuclidianであることとする。これは**エージェントaにとっての度合つき信念**を表す関数

- $\mathcal{A}$の各要素aに対し、Wから$[0,1]$への関数

$$D^a : W \to [0,1]$$

これは**エージェントaにとっての各世界の望ましさ**を定める関数

- 可能世界の集合 W から、2^W への関数

$$\mathcal{G} : W \to 2^W$$

これは以下の制約を受けるものとする。
到達可能性関係 $\mathcal{G}$ と $\mathcal{G}^{-1}$ は、

– 反射的
– 推移的
– $w_1, w_2 \in \mathcal{G}(w)$ ならば、$w_1 \in \mathcal{G}(w_2)$ もしくは $w_2 \in \mathcal{G}(w_1)$

これは可能世界 w の未来の可能世界 $\mathcal{G}(w)$ を表す関数

- $\mathcal{A}$の各要素aに対し、$w' \in \mathcal{G}$を満たすwとw'の組から、$[0,1]$への関数

$$E^a : \{(w, w') \mid w' \in \mathcal{G}(w)\} \to [0,1]$$

これは**エージェントaの努力の度合い**を表す関数

以上を組にしたもの

$$M = \langle W, V, \{R^a_B \mid a \in \mathcal{A}\}, \{D^a \mid a \in \mathcal{A}\},$$
$$\{\mathcal{G} \mid w \in W\}, \{E^a \mid a \in \mathcal{A}\}\rangle$$

を構造と呼ぶ。

解釈

- $p \in S$ ならば、
  $[\![p]\!]_{\langle M,w \rangle} = \top$ **iff** $V(w,p) = \top$
- $[\![\neg\phi]\!]_{\langle M,w \rangle} = \top$ **iff not** $[\![\phi]\!]_{\langle M,w \rangle} = \top$
- $[\![\phi \vee \psi]\!]_{\langle M,w \rangle} = \top$ **iff** $[\![\phi]\!]_{\langle M,w \rangle} = \top$ **or** $[\![\psi]\!]_{\langle M,w \rangle} = \top$

- $[\![Prob^a_\ell \phi]\!]_{\langle M,w \rangle} = \top$ **iff** $\sum_{\substack{w' \in W \text{かつ} \\ w' \text{で} \phi \text{が真}}} R^a_B(w, w') = \ell$

- $[\![G\phi]\!]_{\langle M,w \rangle} = \top$
  iff　全ての$\mathcal{G}(w)$**に対し、**$[\![\phi]\!]_{\langle M,w' \rangle} = \top$

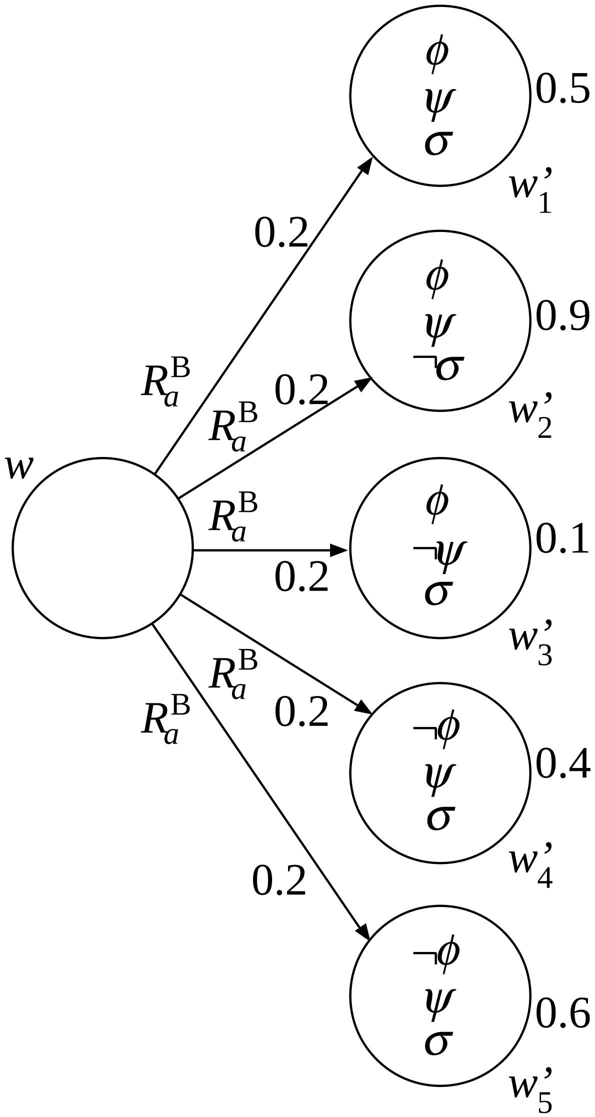
- $[\![H\phi]\!]_{\langle M,w \rangle} = \top$
  iff $w \in \mathcal{G}(w')$ **を満たす全ての**w'**に対し、**
  $[\![\phi]\!]_{\langle M,w' \rangle} = \top$

- $[\![P\phi]\!]_{\langle M,w\rangle} = \top$
  iff $w \in \mathcal{G}(w')$ **を満たす** w' **が存在して、** $[\![\phi]\!]_{\langle M,w'\rangle} = \top$

- $[\![P^a_e\phi]\!]_{\langle M,w\rangle} = \top$
  iff $w \in \mathcal{G}(w')$ **を満たす** w' **が存在して、**
  $[\![\phi]\!]_{\langle M,w'\rangle} = \top$ **and** $E^a(w, w') = e$

- $[\![Des^a_{=d}\phi]\!]_{\langle M,w\rangle} = \top$ **iff** $\dfrac{\sum_{\substack{w'\in W\text{ かつ}\\ w'\text{で}\phi\text{が真}}}(D^a(w')\times R^a_B(w,w'))}{\sum_{\substack{w'\in W\text{ かつ}\\ w'\text{で}\phi\text{が真}}} R^a_B(w,w')} = d$

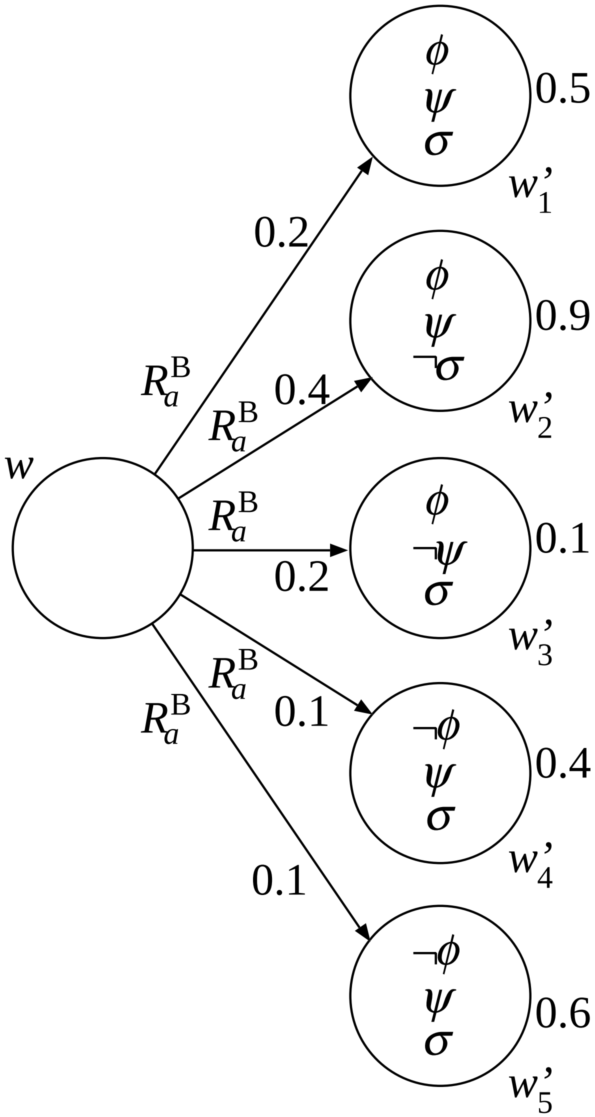
$[\![Des^a_{\geq d}\phi]\!]$などの解釈は、上の式の $= d$を $\geq d$などに変えて定義する



可視関係に付いている値は$R^a_B(w, w')$で、世界に付いている値が$D^a(w')$。
この図では、$[\![Des^a_{=0.5}\phi]\!]_{\langle M,w\rangle} = \top$となる。なぜなら

$$\frac{0.5 \times 0.2 + 0.9 \times 0.2 + 0.1 \times 0.2}{0.2 + 0.2 + 0.2} = 0.5$$

なので。この場合、ϕの望ましさ0.5は、ϕの成り立つ世界3つの望ましさの平均になっている。
同様にして、$[\![Des^a_{=0.7}(\phi \wedge \psi)]\!]_{\langle M,w\rangle} = \top$, $[\![Des^a_{=0.3}(\phi \wedge \sigma)]\!]_{\langle M,w\rangle} = \top$ となる。



もし各可視関係に付いている値が上から $0.2,\ 0.4,\ 0.2,\ 0.1,\ 0.1$ だったら、ϕ の望ましさは

$$\frac{0.5 \times 0.2 + 0.9 \times 0.4 + 0.1 \times 0.2}{0.2 + 0.4 + 0.2} = 0.6$$

となり、w'_2 の望ましさに引きずられて高くなる。このように、ϕ の望ましさは、ϕ の成り立つ世界 3 つの望ましさの加重平均になる。

この意味論は以下の性質を満たす。

1. $\phi \Leftrightarrow \psi$が恒真ならば、任意の$M$, w, a, dに対し $[\![Des^{a}_{=d}\phi]\!]_{\langle M,w\rangle} = \top$ **iff** $[\![Des^{a}_{=d}\psi]\!]_{\langle M,w\rangle} = \top$

ϕとψが同値ならば、それらの望ましさは等しくあるべき

この意味論は以下の性質を満たす。

2. $[\![Des^a_{=d_1}\phi]\!]_{\langle M,w\rangle} = \top$ かつ $[\![Des^a_{=d_2}(\phi \wedge \psi)]\!]_{\langle M,w\rangle} = \top$ のとき、d_2はd_1より大きいことも小さいこともある。すなわち、ϕより$\phi \wedge \psi$の方が望ましいことも、望ましくないこともある。

例えば、「単位がとれる」よりも
「"単位がとれる" かつ "卒業できる"」は望ましいが、
「"単位がとれる" かつ "留年する"」は望ましくない。

ここでさらに、R^a_Bに以下の制限を入れる。これは、普通の信念の論理において、信念を表す世界間の可視関係に、推移性とEuclidean性を入れることに相当する。

- $R^a_B(w, w') > 0$となるような任意の$w, w' \in W$、および任意の$w'' \in W$に対し、$R^a_B(w, w'') = R^a_B(w', w'')$

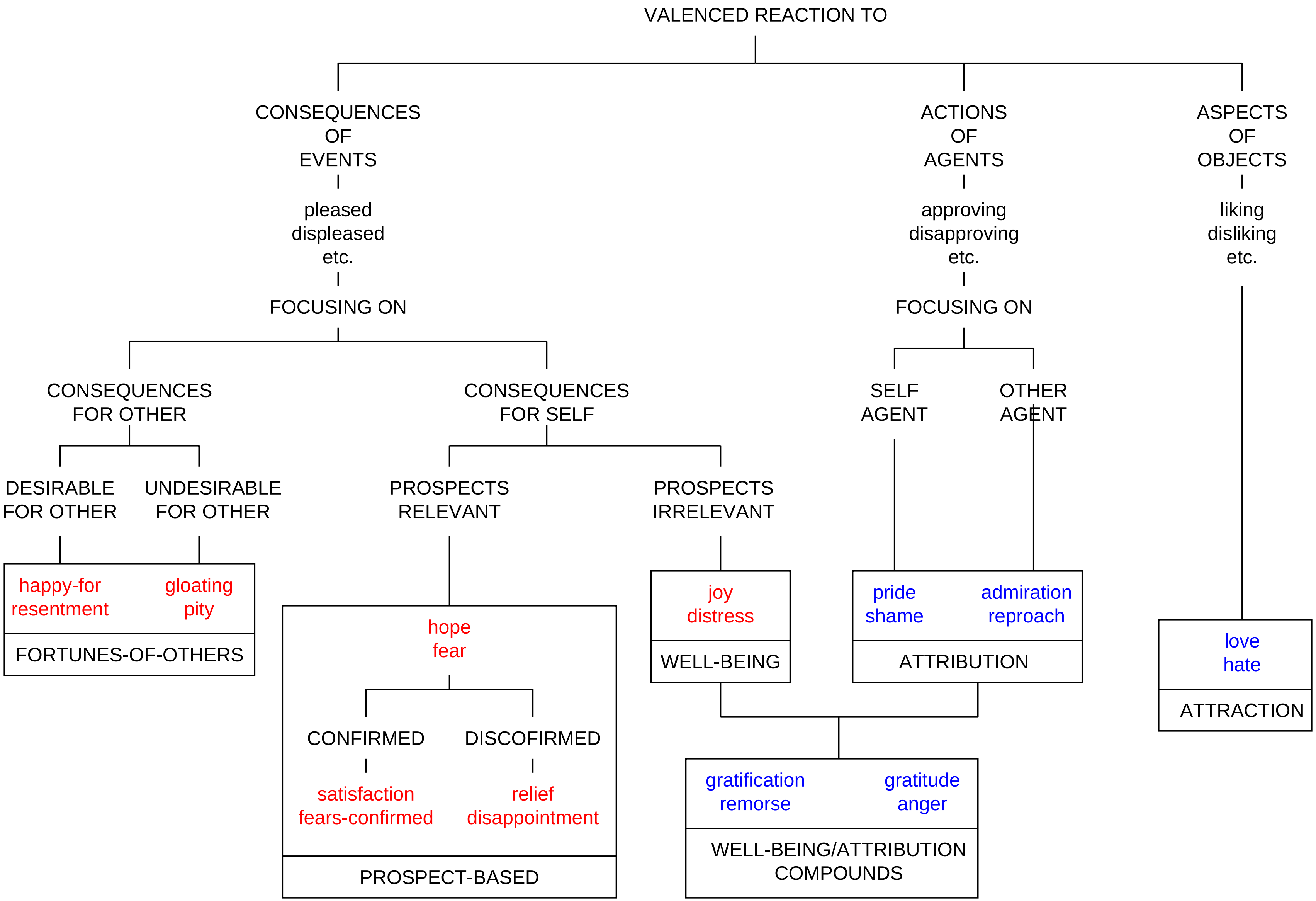
この制約を入れることで、以下の性質が成り立つ。

- $Prob^a_\ell \supset Bel^a Prob^a_\ell \phi$ や $Bel^a Prob^a_\ell \phi \supset Prob^a_\ell$ は恒真
- その特別な場合として、
  $Bel^a \supset Bel^a Bel^a \phi$ や $Bel^a Bel^a \phi \supset Bel^a$ は恒真
- $Des^a_{=d}\phi \supset Bel^a Des^a_{=d}\phi$ や $Bel^a Des^a_{=d}\phi \supset Des^a_{=d}\phi$ も恒真
- Joy の定義を用いると、
  $Joy^a_d\phi \supset Bel^a Joy^a_d\phi$ や $Bel^a Joy^a_d\phi \supset Joy^a_d\phi$ も恒真

このように、自分の信念や感情に関して自分が正しい信念を持っているという、望ましい性質が出てくる。

以上の論理体系で、OCC Theoryで定義されている、22種類の感情のうち、以下の12種類の度合い付き表現が可能に。

- 喜び(Joy)
- 嘆き(Distress)
- 共に喜ぶ(Happy-For)
- 共に残念に思う(Sorry-For)
- 憤り(Resentment)
- 嘲笑う(Gloating)
- 望み(Hope)
- 恐れ(Fear)
- 安堵(Relief)
- 落胆(Disappointment)
- 満足(Satisfaction)
- 恐れていたことが現実になる(Fear-Confirmed)


VALENCED REACTION TO
CONSEQUENCES OF EVENTS
pleased displeased etc.
FOCUSING ON
CONSEQUENCES FOR OTHER
DESIRABLE FOR OTHER
UNDESIRABLE FOR OTHER
happy-for resentment
gloating pity
FORTUNES-OF-OTHERS
CONSEQUENCES FOR SELF
PROSPECTS RELEVANT
hope fear
CONFIRMED
DISCOFIRMED
satisfaction fears-confirmed
relief disappointment
PROSPECT-BASED
PROSPECTS IRRELEVANT
joy distress
WELL-BEING
ACTIONS OF AGENTS
approving disapproving etc.
FOCUSING ON
SELF AGENT
OTHER AGENT
pride shame
admiration reproach
ATTRIBUTION
gratification remorse
gratitude anger
WELL-BEING/ATTRIBUTION COMPOUNDS
ASPECTS OF OBJECTS
liking disliking etc.
love hate
ATTRACTION

- **喜び** (Joy)
  $Bel^a \phi \wedge Des^a_{=d} \phi \supset Joy^a_d \phi$
- **嘆き** (Distress)
  $Bel^a \phi \wedge Des^a_{=d} \neg \phi \supset Distress^a_d \phi$

- **共に喜ぶ**(Happy-For)

$$Bel^a \phi \wedge Bel^a Des^b_{=d_1} \phi \wedge Des^a_{=d_2} Bel^b \phi$$
$$\supset HappyFor^{a,b}_{d_1 \times d_2} \phi$$

- **共に残念に思う**(Sorry-For)

$$Bel^a \phi \wedge Bel^a Des^b_{=d_1} \neg \phi \wedge Des^a_{=d_2} \neg Bel^b \phi$$
$$\supset SorryFor^{a,b}_{d_1 \times d_2} \phi$$

- **憤り** (Resentment)

$$Bel^a\phi \wedge Bel^a Des^b_{=d_1}\phi \wedge Des^a_{=d_2}\neg Bel^b\phi \supset Resentment^{a,b}_{d_1\times d_2}\phi$$

- **嘲笑う** (Gloating)

$$Bel^a\phi \wedge Bel^a Des^b_{=d_1}\neg\phi \wedge Des^a_{=d_2} Bel^b\phi \supset Gloating^{a,b}_{d_1\times d_2}\phi$$

- **望み**(Hope)
  $$Prob^a_\ell \phi \wedge Des^a_{=d} \phi \wedge \ell \neq 1 \supset Hope^a_{d \times (1-\ell)} \phi$$
- **恐れ**(Fear)
  $$Prob^a_\ell \phi \wedge Des^a_{=d} \neg \phi \wedge \ell \neq 1 \supset Fear^a_{d \times \ell} \phi$$

- **満足**(Satisfaction)

$$Des^a_{=d}\phi \wedge Bel^a\phi \wedge Bel^a P^a_e Prob^a_\ell \phi \wedge \ell \neq 1$$
$$\supset Satisfaction^a_{d\times\ell\times e}\phi$$

- **恐れていたことが現実になる**(Fear-Confirmed)

$$Des^a_{=d}\neg\phi \wedge Bel^a\phi \wedge Bel^a P^a_e Prob^a_\ell \phi \wedge \ell \neq 1$$
$$\supset FearConfirmed^a_{d\times\ell\times e}\phi$$

- **安堵**(Relief)

$$Des^a_{=d}\phi \wedge Bel^a\phi \wedge Bel^a P^a_e Prob^a_\ell \neg\phi \wedge \ell \neq 1 \\ \supset Relief^a_{d\times\ell\times e}\phi$$

- **落胆**(Disappointment)

$$Des^a_{=d}\neg\phi \wedge Bel^a\phi \wedge Bel^a P^a_e Prob^a_\ell \neg\phi \wedge \ell \neq 1 \\ \supset Disappointment^a_{d\times\ell\times e}\phi$$

課題

- OCC Theoryで定義されている、残りの感情の度合い付き表現を可能にするため、新たに変数を導入し、論理モデルを拡張する
- 導入した変数d, ℓ, eを定めるの関数のもっともらしさを検討する必要がある
- 実際に実装が可能な論理モデルであるかを検討する必要がある