

Émotions artificielles et processus cognitifs de résilience

Clément Raïevsky

Conduite des Systèmes Dynamiques & IA

7 Novembre 2019



LCIS - MACSY-COSY

Approches complémentaires :

- Automatique - Théorie du contrôle
- Intelligence Artificielle Collective

Modélisation, commande et supervision des systèmes complexes artificiels

Plan

1. Contexte

2. Résilience

3. Théories psychologiques des émotions

4. Détection des situations particulières - Génération des émotions

5. Réaction à une situation particulière - Influence des émotions

6. Adaptations

7. Conclusion - Perspectives

Contexte

Contexte

Objectif

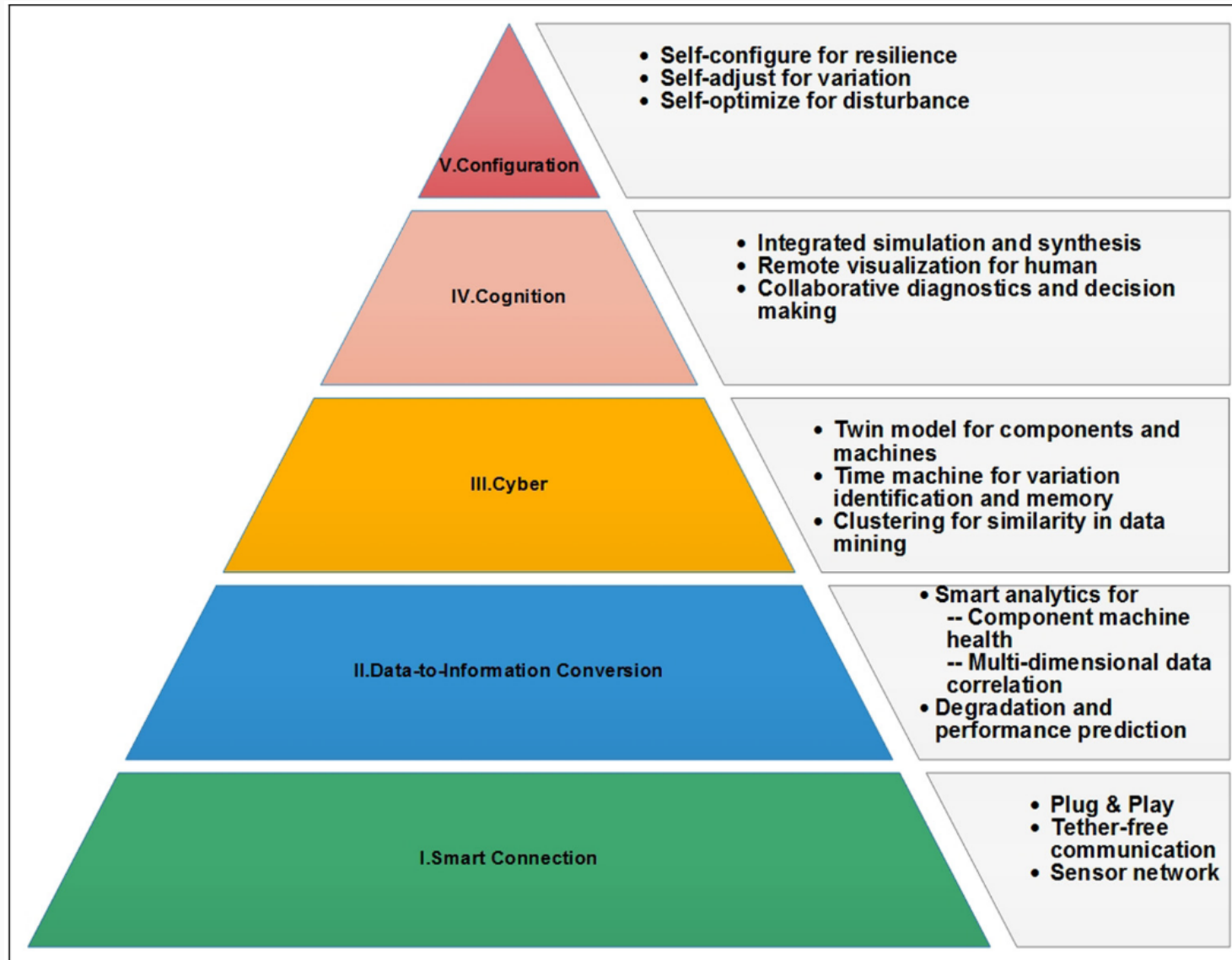
Augmenter la **résilience** des systèmes cyber-physiques (CPS)

Approche

Intelligence Artificielle Collective - Systèmes multi-agents (SMA)

- Systèmes ouverts - distribués
- Connexion non garantie
- Nécessité d'une forme d'**autonomie de décision**

Positionnement dans l'architecture 5C



Résilience

Résilience

Définition adoptée

Capacité d'un système à :

- Détecter
- Réagir
- S'adapter

aux situations ou évènements particuliers

Détection des situations particulières

Caractère **particulier**

Comportements, capacités courantes non adaptées

Dépend de la capacité du concepteur à prévoir les situations rencontrées

- Forcément limitée,
- D'autant plus dans les CPS car ils sont :
 - Ouverts
 - Dynamiques

Réaction

Comportements visant à **conserver** le fonctionnement du système

- Modification du :
 - choix du comportement
 - mode de fonctionnement des comportements
- Application de plans prédéfinis
- Perception ou communication en vue de confirmer le caractère particulier de la situation
- Identification de la cause de la situation (diagnostic)
- Isolation des composants impliqués
- . . .

Adaptation

Modifications des comportements en vue d'améliorer

- la **détection** de situations similaires
- la **réaction** à ces situations

Comportements Individuels

- Apprentissage
- Priorités

Modification de l'organisation

- Topologie
- Confiance

Quel rapport avec les émotions ?

**Les émotions sont des processus cognitifs
cruciaux pour notre résilience**

Théories psychologiques des émotions

Théories psychologiques des émotions

Pas de théorie unifiée

- Une multitude de définitions
- Des termes recouvrant des phénomènes très différents
 - Surprise
 - Honte
 - Amour
- Le concept d'émotion est même parfois remis en cause (Griffith, Duffy)

Consensus

Les émotions jouent un rôle fondamental dans notre cognition.

Consensus sur les émotions

Fonctions pour l'individu :

- Interrompre le fonctionnement courant de la prise de décision
- Focaliser l'attention sur la cause de l'émotion
- Signaler les situations particulières

Fonctions pour le groupe :

- Assurer une forme de cohérence entre les membres du groupe
- Établir et maintenir les liens entre les membres du groupe (engagements, structuration sociale)

Fonctions des émotions relatives à la résilience

Détection

Processus cognitif incrémental

- Réponse physiologique
- Influence sur la cognition
- Comportements dédiés à :
 - confirmer le caractère particulier
 - modifier la préparation à l'action
- Déclenche la phase de réaction

Influence collective

- Perception des émotions des autres influence notre détection

Fonctions des émotions pour la résilience

Réaction

Réaction individuelle

- Modifier la préparation à l'action
- Réguler le niveau de dépense énergétique
- Déclencher des comportements visant à confirmer le caractère particulier de la situation
- Focaliser l'attention sur la cause de la situation

Réaction collective

- Exprimer l'occurrence d'un épisode émotionnel
- Déclencher une émotion à partir de l'état des autres

Fonctions des émotions pour la résilience

Adaptation

Individu

- Influencer l'apprentissage relatif à la situation courante

Collectif

- Modifier les relations au sein du groupe
- Réguler les engagements

Construire des processus de prise de décision artificiels qui reproduisent les fonctions cognitives des émotions relatives à la résilience

Émotions Artificielles

Principaux domaines d'application :

- Interaction Homme-Machine
 - Agents conversationnels
 - Believable Agents
 - Reconnaissance des émotions humaines
- Simulation sociale

Notre proposition

Reproduire les fonctions cognitives des émotions plutôt que de simuler leurs effets

Processus cognitifs de résilience

En mesure de :

- Détecter
- Réagir
- S'adapter

situations ou évènements particuliers

Détection des situations particulières - Génération des émotions

Génération d'une émotion

Comment initier un épisode émotionnel ?

Les théories psychologiques des émotions :

- Se basent la plupart du temps sur des informations non disponibles aux agents situés :
 - Caractère plaisant/déplaisant de la situation
 - Influence positive/négative sur les préoccupations
- Se focalisent sur la différenciation de l'émotion :
 - Quelle émotion dans quelle situation ?

On retrouve ces caractéristiques dans la plupart des travaux existants

Le **déclenchement** de l'émotion reste peu traité

Génération d'une émotion

Comment initier un épisode émotionnel ?

Pas de théorie ou modèle transposable directement aux systèmes artificiels → Choix d'une classe de théories plutôt que d'une en particulier

Classification des théories par Frijda

- Stimulus particulier
- Intensité
- **Concordance** (*Match-Mismatch*)

Théories de la concordance

Match-Mismatch Theories

Émotions déclenchées par l'apparition
d'une **disparité** ou d'une **adéquation** entre :

- les préoccupations de l'individu et
- la situation courante

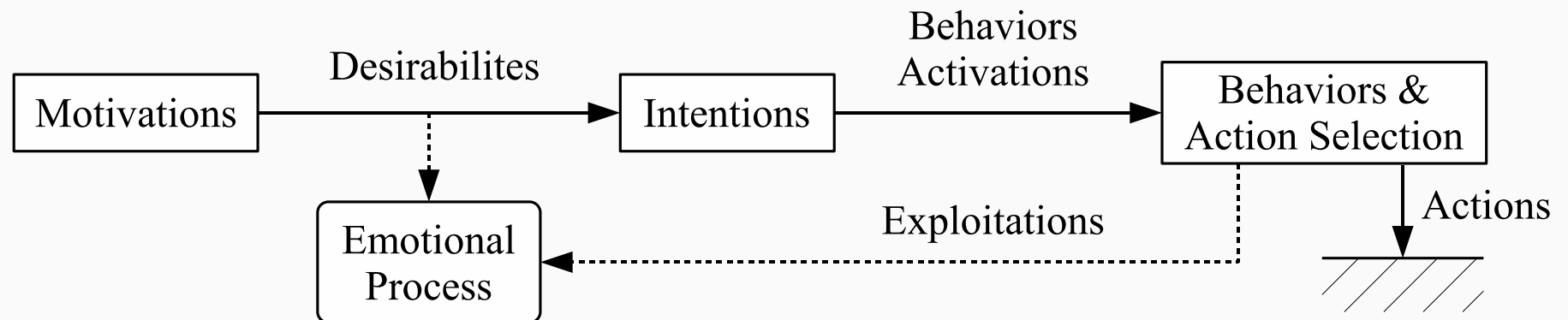
Les préoccupations recouvrent :

- les buts
- les attentes
- les intentions
- les tendances à l'action
- les valeurs
- les engagements

Mécanisme de prise de décision

Comment transposer la théorie de la concordance aux système artificiels ?

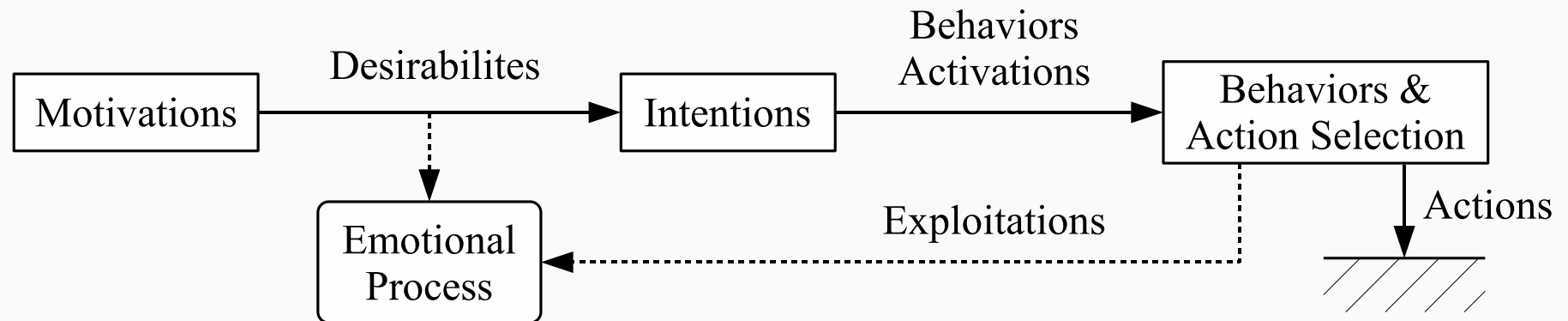
Architecture de prise de décision, version minimaliste



Mécanisme de prise de décision

Comment transposer la théorie de la concordance aux système artificiels ?

Architecture de prise de décision, version minimaliste



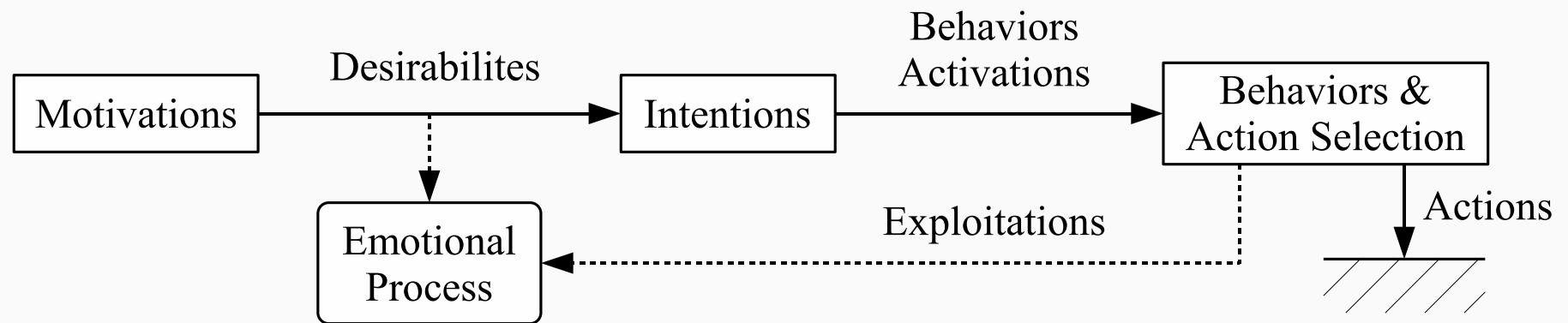
Motivations

- Génèrent des intentions
- Formulent des désirabilités

Mécanisme de prise de décision

Comment transposer la théorie de la concordance aux systèmes artificiels ?

Architecture de prise de décision, version minimaliste



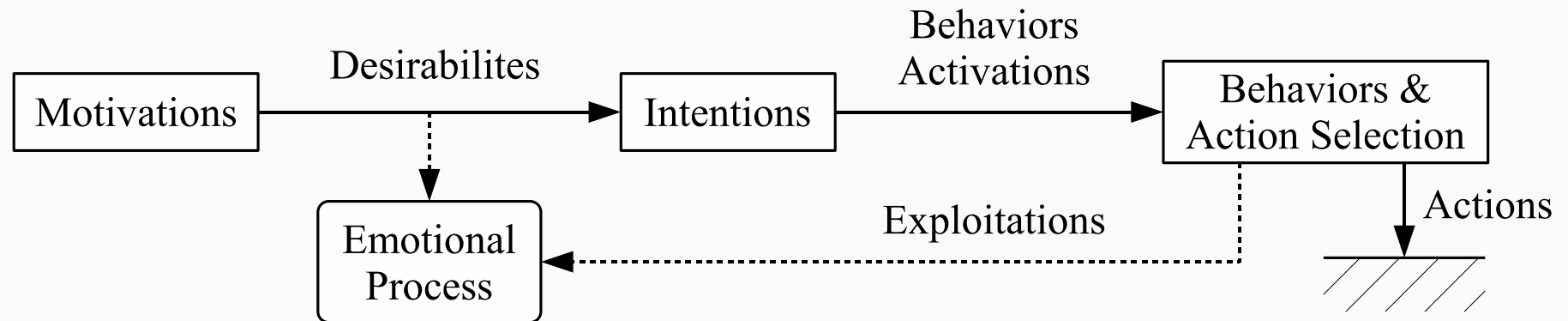
Intentions

- Activent des comportements
- Sont "exploitées" quand les actions de l'agent correspondent à ces comportements

Mécanisme de prise de décision

Comment transposer la théorie de la concordance aux système artificiels ?

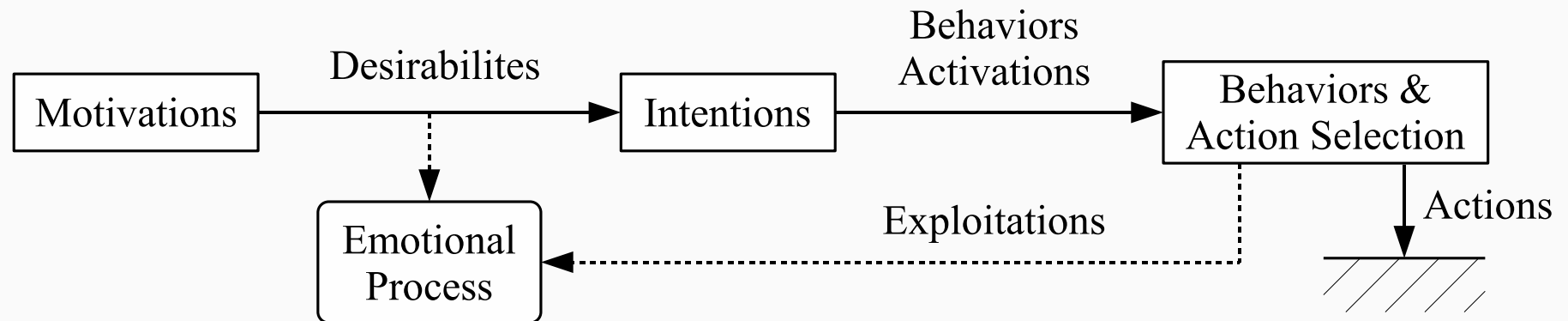
Architecture de prise de décision, version minimaliste



Actions effectives déterminées par :

- La situation
- Les comportements activés

Déclenchement d'une émotion négative

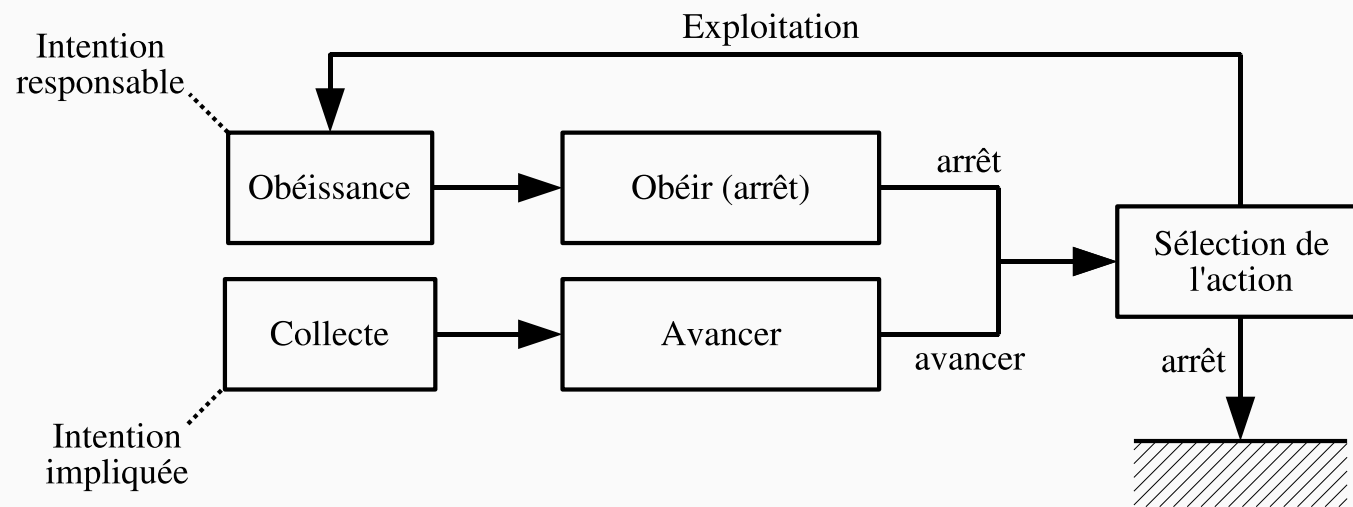
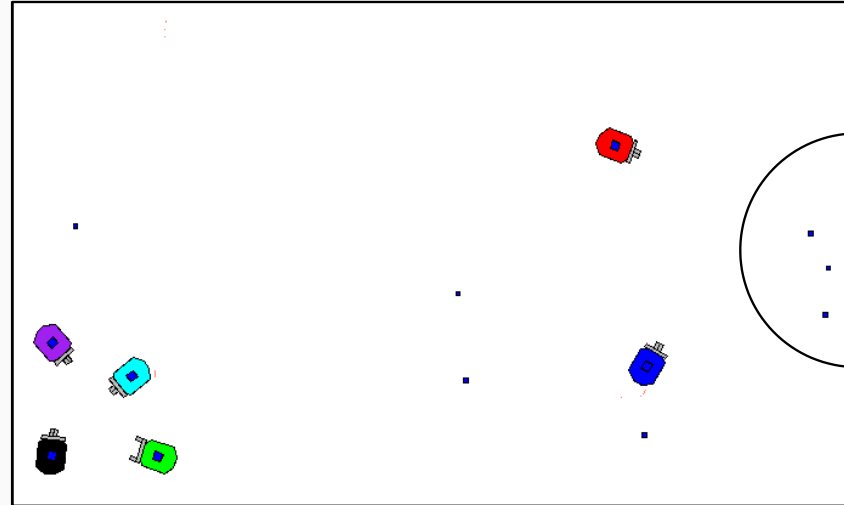


Déclenchement d'une émotion lorsque qu'une **discordance** apparaît entre :

- les intentions et
- la situation

Discordance détectée grâce à l'observation de la **satisfaction des intentions**

Exemple



Réaction à une situation particulière - Influence des émotions

Influences des émotions

Conséquences de la détection d'une disparité

Modification de l'état émotionnel courant

- En fonction du type d'intention impliquée dans la disparité

Mise en valeur de la cause de la disparité

- À partir d'une analyse des intentions courantes
- L'intention responsable est **signalée** aux motivations

Ce sont les *motivations* qui adaptent le comportement de l'agent

Modification de l'état émotionnel

(1/2)

Différentiation entre la peur et colère

Lors d'une discordance avec une préoccupation :

- la **peur** est causée par une perte de contrôle (Lazarus, Frijda)
- la **colère** si l'agent a encore une capacité d'action sur son environnement

Lorsqu'une intention relative à la **sécurité** a déclenché l'émotion

- L'intention est exploitée → les comportements associés sont utilisés mais n'ont pas l'effet attendu
- Perte de contrôle → **Peur**

Modification de l'état émotionnel

(2/2)

Différentiation entre la peur et colère

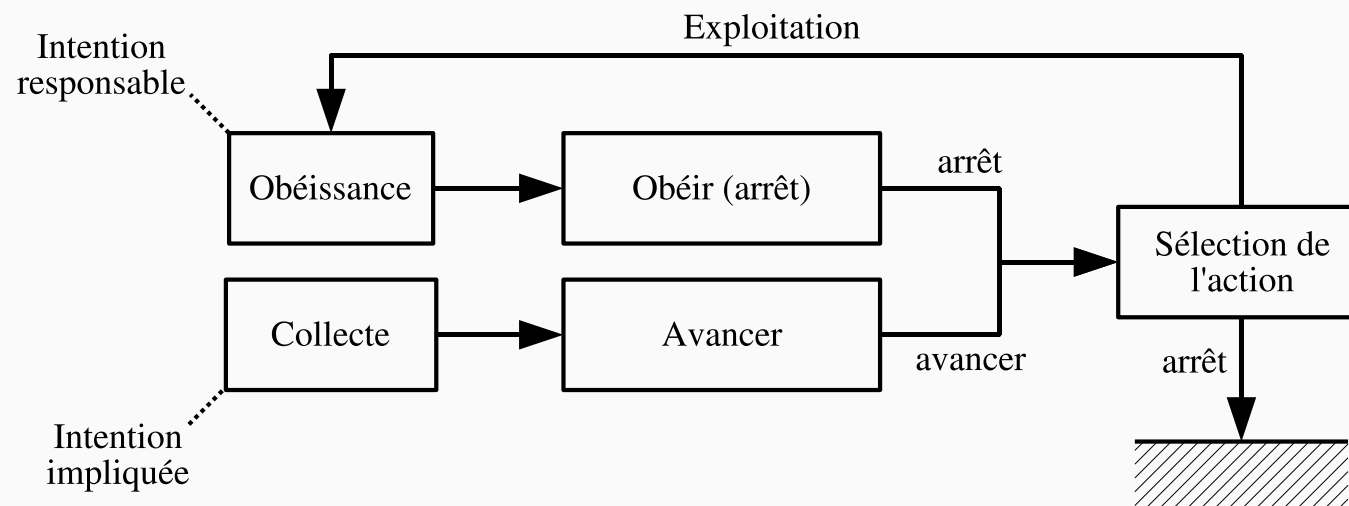
Lorsqu'une intention relative à un **objectif** a déclenché l'émotion

- L'intention n'est pas exploitée
- Les comportements associés ne sont pas utilisés
- Pas de perte de contrôle
- IRO → **Colère**

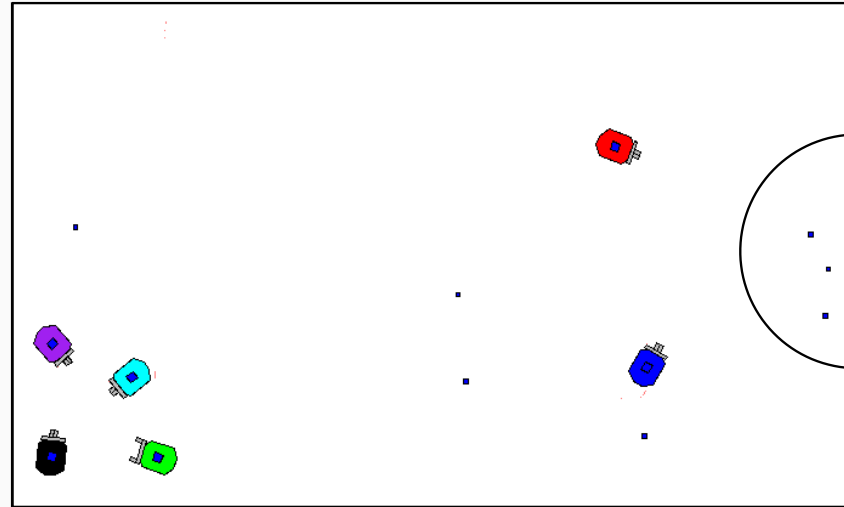
Mise en valeur de la cause de la disparité

Recherche de l'intention responsable

- L'intention qui brise son modèle temporel est l'intention **impliquée**
- L'intention qui à causé ce bris est l'intention **responsable**
- Analyse des intentions utilisées dans le passé proche
- En fonction du type de l'intention impliquée



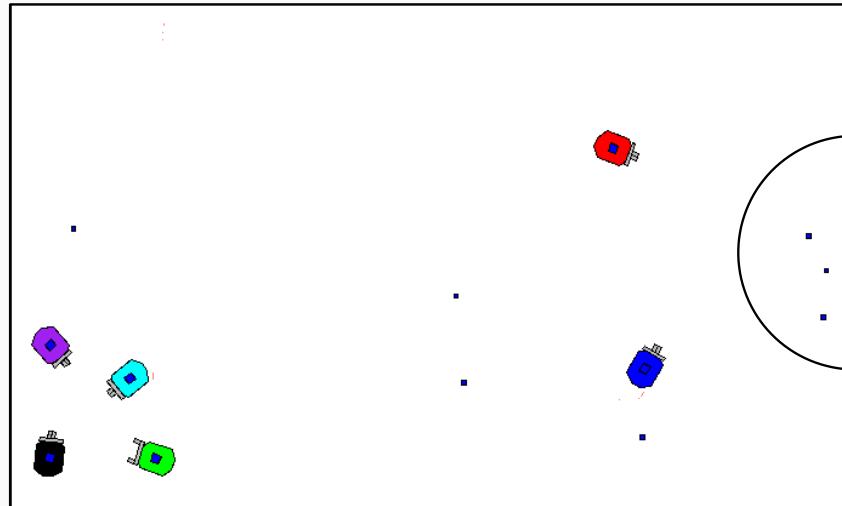
Exemples de réactions



Modification des paramètres des comportements en fonction de l'état émotionnel courant

- Vitesse de déplacement
- Tolérance sur la proximité des obstacles
- Distance d'arrêt devant un supérieur

Exemples de réactions

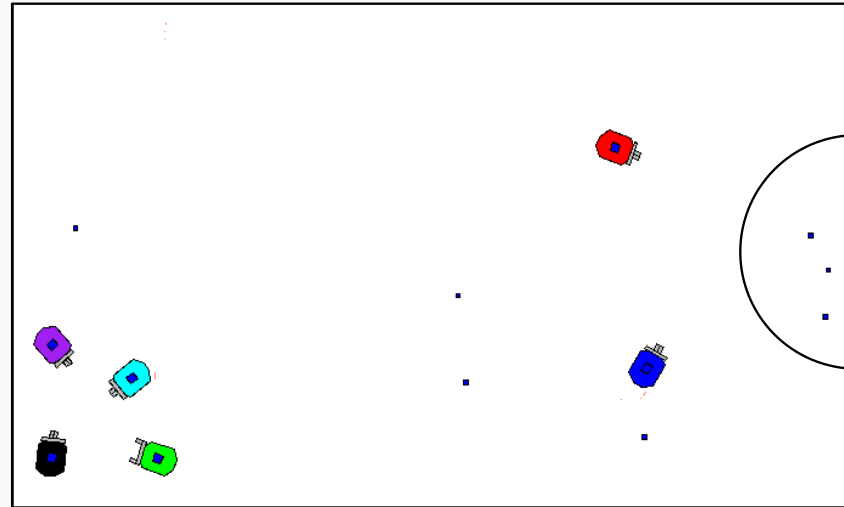


Déclenchement de comportements spécifiques

- Fuite

Adaptations

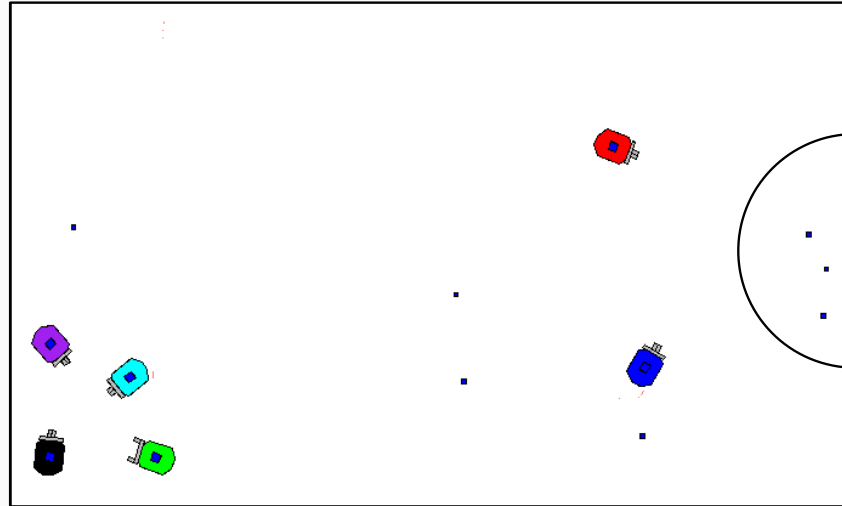
Exemples d'adaptations



Modification des paramètres de détection

- Seuils de déclenchement
- Tolérance de la proximité des obstacles
- Distance d'arrêt devant un supérieur

Exemples d'adaptations



Modification de l'organisation

- Remise en cause de la hiérarchie

Conclusion - Perspectives

Perspectives

Détection

- Mécanismes génériques et incrémentaux de détection d'anomalies dans les données capteurs
- Analyses statistiques simples
- Apprentissage

Perspectives

Réaction

- Perception active
- Coopération pour confirmer le caractère particulier des situations

Adaptation

- Modifications drastiques de l'organisation sociale
 - Ré-organisation
 - Auto-organisation
- Déclenchement de phases d'apprentissage

Conclusion

La résilience consiste à

- Détecter, Réagir et s'adapter
- aux situations particulières

Les émotions jouent un rôle fondamental dans notre résilience

- signalent les situations particulières
- modifient notre préparation à l'action
- influencent l'adaptation de notre comportement

Reproduire les fonctions cognitives de résilience des émotions dans les systèmes artificiels

Merci de votre attention

Généricité de l'influence du processus émotionnel

Les influences du processus émotionnel

- Ne dépendent pas
 - d'un ensemble de situations prédéfinies (indépendant des perceptions)
 - de la mission des agents (indépendant des comportements)
- ne déclenche pas directement de comportements
- remplit la fonction de signal des émotions

Les contraintes posées pour qualifier un processus cognitif d'émotionnel sont respectées

Généricité du modèle

Indépendance du modèle vis-à-vis de :

- un ensemble de situations prédéfinies (indépendant des perceptions)
- la mission des agents (indépendant des comportements)

Limitation et identification des *connaissances introduites par le concepteur*

- Un concepteur ne peut pas savoir comment un agent expérimente la réalité (Nagel)
- Les agents situés évoluent dans un environnement imprévisible

Satisfaction des intentions

Intentions de deux natures différentes

- Intentions relatives à la sécurité (IRS)
- Intentions relatives à un objectif (IRO)

Une Intention relative à la sécurité est satisfaite si :

Elle n'est pas exploitée → les comportements associés ne sont pas utilisés

Une Intention relative à un objectif est satisfaite si :

Elle est exploitée lorsque désirable
→ les comportements associés sont utilisés

Modèle temporels des intentions (2/2)

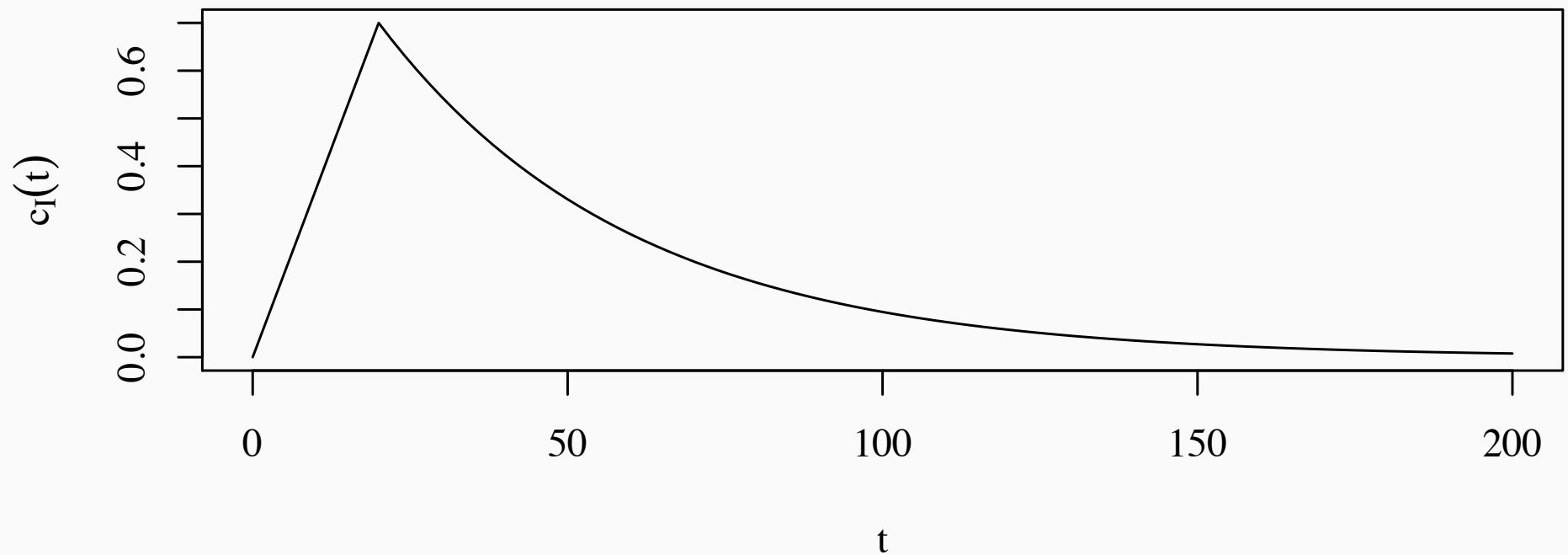
- Mesure du temps cumulé $a_l(t)$ durant lequel les intentions ne respectent pas leurs condition de satisfaction
- Sur une fenêtre de temps glissante de durée F_l
- **Disparité** quand $a_l(t)$ dépasse un seuil S_l

$$a_l(t) = \int_{-F_l}^0 b_l(t) dt \quad (1)$$

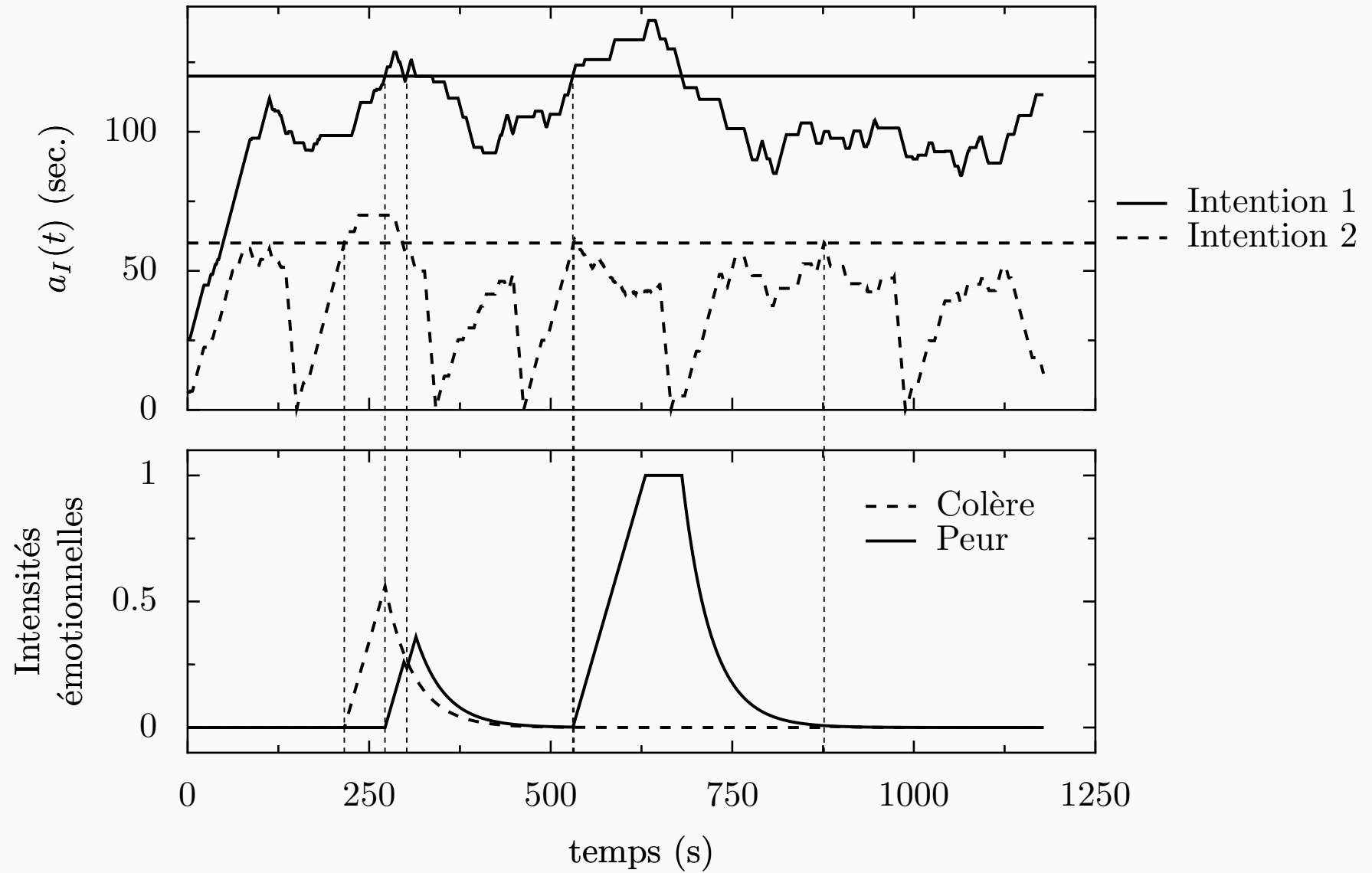
$$b_l(t) = \begin{cases} 0 & \text{si l'intention } l \text{ respecte sa condi-} \\ & \text{tion de satisfaction au temps } t \\ 1 & \text{sinon} \end{cases} \quad (2)$$

Comportement des variables émotionnelles

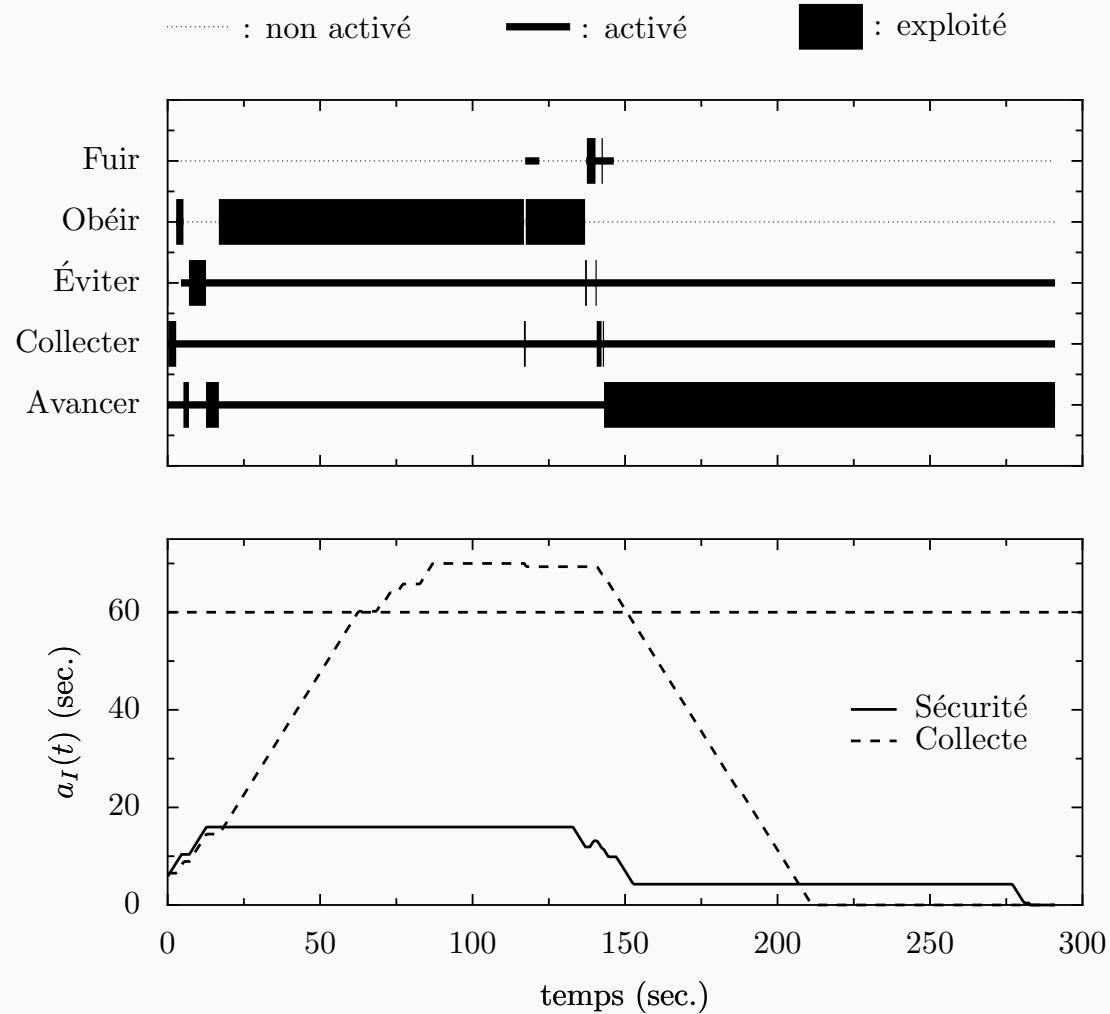
$$c_I(t) = \begin{cases} c_I(t_0) + A_I \cdot (t - t_0) & \text{si } a_I(t) > S_I \\ c_I(t_1) \cdot e^{-D_I \cdot (t - t_1)} & \text{sinon} \end{cases} \quad (3)$$



Exemple



Évolution des variables au cours d'une situation problématique



Architecture de prise de décision utilisée pour les expériences

