A Tuning Machine for Collective Commitments

Barbara Dunin-Kęplicz
Institute of Informatics
Warsaw University and
Institute of Computer Science
Polish Academy of Sciences
and
Rineke Verbrugge
Institute of Artificial Intelligence
University of Groningen

Definition of agent

An *agent* is a computational entity (program or robot) *situated* in some environment, that is capable of *flexible autonomous* action in order to meet its design objectives.

Situatedness: the agent receives sensory input from its environment and can perform actions which change it.

Autonomy: the system should be able to act without the direct intervention of humans (or other agents), and have control over its own actions and internal state.

Definition of agent

Flexibility, amounts to the agent being:

- *responsive*: responding in a timely fashion to changes in the environment ;
- *pro-active*: exhibiting opportunistic, goal-directed behaviour and taking initiative;
- *social*: interacting, when appropriate, with other artificial agents and humans in order to complete their own problem solving and to help others with their activities.

Multi-agent Systems

Multi-agent Systems (MAS) are computational systems in which a collection of loosely coupled *autonomous agents* interact in order to solve a given problem, which is usually beyond the agents' individual capabilities.

MAS represent a new way of:

- analyzing
- designing
- implementing

complex software systems.

The agent-based view offers a powerful repertoire of tools, techniques and metaphors that improve the way in which people conceptualize and implement many types of software.

Multi-agent Systems

MAS usually work in a dynamic and often unpredictable environment:
conditions too hard for conventional knowledge-based systems.

Variety is the core of multi-agent systems: developed

- along different dimensions
- according to various problem solving paradigms

Narrowing the perspective, we will focus on **cooperative problem solving (CPS)**.

Cooperative Problem Solving

Essential aspects of CPS are:

- **cooperation**: working together towards a common aim

- **coordination**: organising problem solving so that harmful interactions are avoided and beneficial ones exploited

- **negotiation**: in the presence of conflict, coming to an agreement which is acceptable to all the parties involved

For all aspects, communication is necessary.

Communication versus reasoning

Belief Desire Intention (BDI) Architectures

Practical reasoning intensional systems

BDI-agents are characterised by "mental state" :

- **beliefs** - information the agent has about its environment

- **goals** - "options" available to the agent, i.e. different states of affairs that the agent may choose to commit to

- **intentions** - states of affairs that the agent has chosen

- in our approach, intentions are viewed as an inspiration for a goal-directed activity, reflected in **commitments**

Practical reasoning

An agent's practical reasoning involves:

1. repeatedly updating beliefs about the environment

2. deciding what options are available

3. "filtering" these options to determine new intentions

4. creating commitments on the basis of intentions

5. performing actions in accordance with commitments.

BDI usually provides an explicit model of *teamwork*.

Teamwork

Teamwork is vital in Multi-agent systems.

In teamwork:

1. Actions of group members are coordinated; but also

2. The group acts as a single agent with beliefs, intentions and commitments of its own, over and above individual ones.

Specifically, a team *collectively intends* to fulfill a clearly defined overall goal φ.

Example:
The Apollo Project, with $\varphi =$ "to place a man on the moon by the end of the 1960s".

Modal operators for group beliefs

BEL(i, φ): agent i believes φ

E-BEL$_G(\varphi)$: each agent in group G believes φ

Collective belief C-BEL$_G(\varphi)$:
everyone in G believes φ, everyone in G believes that everyone in G believes φ, etc.

Axioms:
$KD45_n^C$ is the modal system $KD45_n$ plus:

E-BEL$_G(\varphi) \leftrightarrow \bigwedge_{i \in G}$ BEL(i, φ)

C-BEL$_G(\varphi) \rightarrow$ E-BEL$_G(\varphi \wedge$ C-BEL$_G(\varphi))$

From $\varphi \rightarrow$ E-BEL$_G(\psi \wedge \varphi)$ infer $\varphi \rightarrow$ C-BEL$_G(\psi)$ (Induction Rule).

Collective beliefs established by
communication + *reasoning*

Different degrees of belief in group

BEL(ψ), E-BEL$_G$(ψ), E-BEL$_G^2$(ψ), ..., C-BEL$_G$(ψ)

For teamwork, E-BEL$_G$(ψ) is not sufficient: C-BEL$_G$(ψ) needed.

Example:
collective actions where each individual success is vital to the result, e.g. coordinated attack, where at starting time needs to be C-BEL$_G$.

Advantage of C-BEL$_G$(ψ):
entails C-BEL$_G$ for logical consequences of ψ. Thus, agents reason in a similar way from ψ and collectively believe in this.

If only E-BEL$_G$(ψ) is established, it is more difficult for agents to maintain a model of others w.r.t. ψ and its consequences.

Establishing E-BEL$_G$(ψ) places less constraints on the communication medium.
C-BEL$_G$(ψ) is easy to understand, hard to achieve.

Individual goals and intentions

Agent's activity starts from **goals**.
As it may have many different objectives, its goals need not be consistent with each other. Then, the agent chooses a limited number of its goals to be **intentions**, in such a way that consistency is preserved.

Individual *goals* and *intentions* are formalized as primitive notions wth a formal semantics.

For the motivational operators GOAL and INT axioms include the basic modal system K_n plus:

$\neg(\text{INT}(i, \psi) \wedge \text{INT}(i, \neg\psi))$ for $i = 1, \ldots, n$

Social commitments

A *social commitment* between two agents is weaker than a collective commitment, but stronger than an individual intention.

If agent i commits to j to do something, then:

- i has the *intention* to do that
- j is *interested* in i fulfilling its intention
- the agents are *collectively aware* about their individual attitudes

$$\begin{aligned} &\text{COMM}(i, j, \alpha) \leftrightarrow \\ &\quad \text{INT}(i, \alpha) \wedge \text{GOAL}(j, done(i, \alpha)) \wedge \\ &\quad \text{C-BEL}_{\{i,j\}}(\text{INT}(i, \alpha) \wedge \text{GOAL}(j, done(i, \alpha))) \end{aligned}$$

Here, $done(i, \alpha)$ means that agent i has just executed action α.

Collective intentions in strictly cooperative groups

φ = the main goal of the system

Conditions for a collective intention:

- All members of the group have the associated *individual intention*
- All members in the group *intend* other members to have the associated individual intention, etc.:
  a mutual intention M-INT$_G(\varphi)$
- It is a collective belief in the group that this mutual intention exists

Axioms for mutual and collective intention

E-INT$_G(\varphi)$: "every agent in G intends φ".

$KD45_n^{\text{M-INT}_G}$ is the modal system $KD45_n$ plus:

E-INT$_G(\varphi) \leftrightarrow \bigwedge_{i \in G}$ INT(i, φ)

M-INT$_G(\varphi) \leftrightarrow$ E-INT$_G(\varphi \wedge$ M-INT$_G(\varphi))$

From $\varphi \rightarrow$ E-INT$_G(\psi \wedge \varphi)$ infer $\varphi \rightarrow$ M-INT$_G(\psi)$ (Induction Rule).

Definition of collective intention:

$$\text{C-INT}_G(\varphi) \leftrightarrow (\text{M-INT}_G(\varphi) \wedge \text{C-BEL}_G(\text{M-INT}_G(\varphi)))$$

Semantics for mutual intentions

- $\text{E-INT}_G^1(\varphi)$ abbreviates $\text{E-INT}_G(\varphi)$
- $\text{E-INT}_G^{k+1}(\varphi)$ abbreviates $\text{E-INT}_G(\text{E-INT}_G^k(\varphi))$

Thus: $w \models \text{M-INT}_G(\varphi)$ iff $w \models \text{E-INT}_G^k(\varphi)$ for all $k \geq 1$.

Define world t to be *G-reachable* from world s if there is a path in the Kripke model from s to t along accessibility arrows I_i associated with members i of G:

$s \models \text{M-INT}_G(\varphi)$ iff $t \models \varphi$ for all t that are G-reachable from s.

The axiom system $KD_n^{\text{M-INT}_G}$ is sound and complete with respect to Kripke models where all n accessibility relations I_i are serial.

Example of a collective intention

Two violinists, a and b, have studied together and toyed with the idea of giving a concert together someday.

Later this becomes more concrete: INT(a, φ) and INT(b, φ), where $\varphi =$ "a and b perform the solo parts of the Bach Double Concerto".

After communicating with each other about this, they start practising together.
A mutual intention M-INT$_G(\varphi)$ is now in place for $G = \{a, b\}$, plus a collective belief about this, so C-INT$_G(\varphi)$.

An opportunity appears: Carnegie Hall plans a concert for Christmas Eve, including the Bach Double Concerto.

Example of a collective intention cntd.

Now a, b refine their collective intention to C-INT$_G(\psi)$, where $\psi =$ "a and b perform the solo parts of the Bach Double Concerto at the Christmas Eve concert in Carnegie Hall".

a, b are chosen to be the soloists, and both sign the appropriate contract.
Because they do this together, they have common knowledge, not merely collective belief, of their mutual intention:
M-INT$_G(\psi) \wedge$ C-KNOW$_G$(M-INT$_G(\psi)$).

Common knowledge can be justified if needed, and a commonly signed contract provides a perfect basis for this.
Thus a, b have developed a very strong variant of collective intention.

Alternative definition of mutual intention

In some situations, e.g. time-critical ones, teamwork may tentatively start even if C-INT$_G(\varphi)$ has not been established, especially in circumstances where it is not possible (yet) to establish a C-BEL$_G$ about M-INT$_G(\varphi)$.

M-INT$'_G(\varphi)$ iff:

- everyone in G intends φ, everyone in G intends that everyone in G intends φ, etc. (as in M-INT$_G(\varphi)$);

- everyone intends that there be collective belief in G of this whole infinite conjunction:
  E-INT$_G$(C-BEL$_G$(M-INT$_G(\varphi)$)).

M-INT$'_G(\varphi) \leftrightarrow$ E-INT$_G(\varphi \wedge$ C-INT$_G(\varphi))$

Alternative definition of collective intention

M-INT$'_G$: a "pre-collective intention", of which the team members will later hopefully establish a collective belief.

An alternative definition of collective intention C-INT$'_G$ based on M-INT$'_G$:

$$\text{C-INT}'_G(\varphi) \leftrightarrow \text{M-INT}'_G(\varphi) \wedge \text{C-BEL}_G(\text{M-INT}'_G(\varphi))$$

M-INT$'_G$ is stronger that M-INT$_G$, but it is not comparable to C-INT$_G$.

C-INT$'_G$ is a stronger notion than C-INT$_G$.

Example of alternative mutual intention

A person c has disappeared under the ice and two potential helpers a and b are near.

a, b do not know each other, and there is no clearly marked initiator.

Communication among them is not possible, e.g. because of strong wind. Limited perception is possible: they can see the other move, but cannot distinguish facial expressions.

Both have the individual intention to help, thus E-BEL$_G(\varphi)$, where $G = a, b$ and $\varphi =$ "c has been rescued".

Two persons are needed for a successful rescue, and this is collectively believed: C-BEL$_G(\psi)$, where $\psi =$ "at least two persons are needed to achieve φ".

Example ... cntd.

As no potential helpers are around, a and b believe that they need to act together. Thus, a mutual intention M-INT$_G(\varphi)$ is established.

a, b may even form individual beliefs about this, i.e. M-INT$_G(\varphi) \wedge$ E-BEL$_G$(M-INT$_G(\varphi)$). However, communication being limited, C-BEL$_G$(M-INT$_G(\varphi)$) cannot be established; so C-INT$_G(\varphi)$ does not hold.

Time is critical, so *some* team-like attitude needs to be established. A collective belief about M-INT$_G(\varphi)$ is necessary to enable co-ordination for mouth-on-mouth breathing and heart massage.

Thus, a, b both *intend* that the C-BEL$_G$ about the mutual intention be established to make teamwork possible, i.e. M-INT$'_G(\varphi)$.

Different kinds of collective intention

One can define various notions of collective intention suitable for specific environments.

Example:
Alternative definition of mutual intention M-INT$'_G$ with corresponding collective intention:
C-INT$'_G(\varphi) \leftrightarrow$ M-INT$'_G(\varphi) \wedge$C-BEL$_G$(M-INT$'_G(\varphi)$)
appropriate for cases in which communicative possibilities are severely limited but collective action is vital.

These cases do not exhaust the possibilities.

Social plan expressions

Teams are created on the basis of *collective intentions*: an inspiration for team activity.

Collective commitment reflects the concrete manner of achieving the main goal φ, based on social plan P.

Individual actions are combined into team actions by social plan expressions:

SP1 If a an individual action and i an agent, then $< a, i >$ is a social plan expression

SP2 If α and β are social plan expressions, then

- $< \alpha ; \beta >$ (sequential composition)
- $< \alpha \parallel \beta >$ (parallellism)

are social plan expressions

Example of a social plan

- Team $G = \{t, l, c\}$
  (theorem prover, lemma prover and proof checker)

- Collective intention in G to prove a new mathematical theorem

- During planning they define two lemmas.
  Correct proofs of these are needed +
  a proof of theorem from their conjunction

One possible social plan P:

$$
\begin{aligned}
&<<<< proveL1, l >; < proveL2, l >>\|\\
&< proveTh, t >>;\\
&<<< checkL1, c >; < checkL2, c >>;\\
&< checkTh, c >>>
\end{aligned}
$$

Correctness of social plans: $constitute(\varphi, P)$

The social plan P should result from the main goal φ by

- task division,
- means-end-analysis, and
- action allocation,

as reflected by the predicate $constitute(\varphi, P)$.

If $constitute(\varphi, P)$ holds, successful realization of the plan P should lead to the achievement of the main goal φ.

Tuning machine for collective commitment

Different notions of collective commitment are sensitive to a variety of circumstances.

Vital ingredients of collective commitment are:

I degrees of belief in a team

II kind of collective intention

III aspects of team awareness

These are viewed as three ‘knobs’, separately tuned in order to obtain appropriate definition of collective commitment for a given context.

General schema for collective commitment

$$
\begin{aligned}
&\text{C-COMM}_{G,P}(\varphi) \leftrightarrow \\
&\quad \text{M-INT}_G(\varphi) \\
&\quad \{\wedge\ awareness_G(\text{M-INT}_G(\varphi))\} \wedge \\
&\quad constitute(\varphi, P) \\
&\quad \{\wedge\ awareness_G(constitute(\varphi, P))\} \wedge \\
&\quad \bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha) \\
&\quad \{\wedge\ awareness_G(\bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha))\}
\end{aligned}
$$

Instantiating the above schema $=$
tuning the $awareness_G$-knobs

- from $\emptyset$
- through individual beliefs and different E-BEL$_G^k$

- to C-BEL$_G$

analogously for degrees of knowledge.

Different aspects of awareness are:

1. mutual intention M-INT$_G$:
   collective awareness of the team
   C-BEL$_G$(M-INT$_G(\varphi)$)

2. correctness of the plan w.r.t. φ:
   collective awareness of this
   C-BEL$_G(constitute(\varphi, P))$

3. Social structure given by social commitments about allocated actions in P:

   (a) collective awareness **de re**, i.e. *detailed* awareness of each social commitment
   $\bigwedge_{\alpha \in P} \bigvee_{i,j \in G}$ C-BEL$_G$(COMM(i, j, α))

   (b) collective awareness **de dicto**, i.e. *global* awareness of the existence of social commitments

   C-BEL$_G$($\bigwedge_{\alpha \in P} \bigvee_{i,j \in G}$ COMM(i, j, α))

The distinction de dicto vs. de re

A sentence of the form $\exists x$BEL$(j, A(x))$ is *de re* belief attribution which relates j to a *res*: individual the belief is about.

BEL$(j, \exists x A(x))$ is a *de dictum* belief attribution, relating j to a *dictum*: proposition $\exists x A(x)$.

Also fruitful for complex epistemic operators such as collective belief.

Note that C-BEL$_G$ in
$\bigwedge_{\alpha \in P} \bigvee_{i,j \in G}$ C-BEL$_G$(COMM(i, j, α)) and

C-BEL$_G$($\bigwedge_{\alpha \in P} \bigvee_{i,j \in G}$ COMM(i, j, α))
distributes over conjunction ($\bigwedge_{\alpha \in P}$), so that only the position of C-BEL$_G$ w.r.t. $\bigvee_{i,j \in G}$ matters.

Different aspects of awareness contd.

When the degree of awareness is expressed in the strongest form of collective belief (C-BEL$_G$), then *aspects of awareness* point out what can be done collectively by the group.

Otherwise, it says nothing explicit about organization of different elements of teamwork, though a matching teamwork organization may be inferred.

Robust collective commitment R-COMM$_{G,P}$

1. Collective intention within the team.

2. Correct plan P leading to φ

3. Collective awareness of correctness of P

4. Social commitments for all actions in P

5. Detailed collective awareness about social commitments

$$
\begin{aligned}
&\text{R-COMM}_{G,P}(\varphi) \leftrightarrow \text{C-INT}_G(\varphi) \wedge \\
&\quad constitute(\varphi, P) \wedge \\
&\quad \text{C-BEL}_G(constitute(\varphi, P)) \wedge \\
&\quad \bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha) \wedge \\
&\quad \bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{C-BEL}_G \text{COMM}(i, j, \alpha))
\end{aligned}
$$

Robust collective commitment cntd.

Example
A company where members are shareholders, including:

- collective responsibility
- collective planning
- collective making of bilateral commitments
- collective revision, if necessary

Strong collective commitment S-COMM$_{G,P}$

1. Collective intention within the team
2. Correct plan P leading to φ
3. Collective awareness of correctness of P
4. Social commitments for all actions in P
5. Global collective awareness about existence of social commitments

$$
\begin{aligned}
&\text{S-COMM}_{G,P}(\varphi) \leftrightarrow \text{C-INT}_G(\varphi) \wedge \\
&\quad constitute(\varphi, P) \wedge \\
&\quad \text{C-BEL}_G(constitute(\varphi, P)) \wedge \\
&\quad \bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha) \wedge \\
&\quad \text{C-BEL}_G\Big(\bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha)\Big)
\end{aligned}
$$

Strong collective commitment cntd.

Example
A company, possibly with leader(s) and rather separate subteams

- collective planning, but

- making of bilateral commitments in relevant subgroups, thus

- no collective responsibility

- no potential of collective revision

Weak collective commitment W-COMM$_{G,P}$

1. Collective intention within the team.

2. Correct plan P leading to φ

3. Social commitments for all actions in P

4. Global collective awareness about existence of social commitments

$$
\begin{aligned}
&\text{W-COMM}_{G,P}(\varphi) \leftrightarrow \text{C-INT}_G(\varphi)\wedge \\
&\quad constitute(\varphi, P) \wedge \\
&\quad \bigwedge_{\alpha\in P} \bigvee_{i,j\in G} \text{COMM}(i,j,\alpha) \wedge \\
&\quad \text{C-BEL}_G\Big(\bigwedge_{\alpha\in P} \bigvee_{i,j\in G} \text{COMM}(i,j,\alpha)\Big)
\end{aligned}
$$

Weak collective commitment cntd.

Example
A company with dedicated planner or planning department

- team knows the overall goal, but
- does not know details of the plan
- it suffices that the planner believes $constitute(\varphi, P)$

Social structure in definitions

These three definitions *implicitly* reflect different social structures in which group is involved.

Social structure originates from different types of hierarchies, based on power relations and dependency relations.

These relations are implicitly reflected in the set of social commitments

$\bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \mathrm{COMM}(i, j, \alpha)$.

Agents' limited orientation in the action distribution may be done on purpose.

Team commitment T-COMM$_{G,P}(\varphi)$

1. Collective intention within the team

2. Correct plan P leading to φ

3. Social commitments for all actions in P

$$
\begin{aligned}
&\text{T-COMM}_{G,P}(\varphi) \leftrightarrow \text{C-INT}_G(\varphi) \wedge \\
&\quad constitute(\varphi, P) \wedge \\
&\quad \wedge \bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha)
\end{aligned}
$$

Team commitment cntd.

Example
A company assigning 'limited trust' to its employees

- team as a structure exists: the overall goal is publicly known, but
- agents remain aware solely about their piece of work, without orientation on the involvement of others: this information may be confidential

Distributed commitment D-COMM$_{G,P}(\varphi)$

1. Correct plan P leading to φ

2. Social commitments for all actions in P

$$\text{D-COMM}_{G,P}(\varphi) \leftrightarrow \mathit{constitute}(\varphi, P) \wedge \\ \wedge \bigwedge_{\alpha \in P} \bigvee_{i,j \in G} \text{COMM}(i, j, \alpha).)$$

A loosely coupled group of agents works in a distributed manner without autonomous involvement in the project.

Distributed commitment cntd.

Example
A company contracting labour to outsiders

- team as a structure does not exist

- agents do not know the overall goal, solely their share in the final project, e.g. in order to avoid potential competition

Awareness of robust collective commitment

$$\text{R-COMM}_{G,P}(\varphi) \rightarrow \text{C-BEL}_G(\text{R-COMM}_{G,P}(\varphi))$$

NOT for weaker cases: agents are not necessarily aware of the strength (kind) of commitment between them.

Intermediate levels of commitment may be characterized by using different degrees of awareness.

Remarks about collective commitment

- The definitions are complex, but not overloaded, and easy to understand and to use.

- Collective commitment may vary in order to adapt to changing circumstances, in such a way that collective intention can still be reached.

- To replan it is often sufficient to revise some of the pairwise social commitments, without the involvement of the entire team.

- When teams social structure changes, this is easy to reflect in a new collective commitment, by changing some of the social ones.

The reconfiguration problem

A dynamic and unpredictable environment poses the problem that team members may fail to realize their tasks or new opportunities may appear.

This leads to the *reconfiguration problem*.

The explicit BDI-model of teamwork helps the team to monitor its performance and to replan based on the present situation.

Advantages of the definitions of collective commitment lead to a solution of **reconfiguration problem**:

(1) the abstract reconfiguration algorithm,

(2) a formal description of cases treated in the algorithm.

Conclusions

- A *static* theory of CPS is built incrementally: starting from individual intentions, through social commitments, leading ultimately to collective intentions and collective commitments.

- All these notions are defined in multi-modal logics with clear semantics.

- Subtle aspects of CPS like various connections between agents may be expressed.

- This diversity of definitions' strength allows to model different situations occurring in practical domains.

Conclusions

- The *dynamic* part of the theory specifies the evolution of motivational attitudes in a dynamic and unpredictable environment.

- The theory may serve a system designer as a specification to create a correct and complete system, as well as to verify it.

- At the IIWU a platform DORCAS enabling a user to build BDI systems containing collective attitudes has been created.

- An interesting instantiation of the reconfiguration algorithm dealing with emergency situations on a boat is implemented.

Selected bibliography

B. Dunin-Kęplicz and R. Verbrugge,
A reconfiguration algorithm for distributed problem solving, *Electronic Modeling* 22 (2000), 68-86

F. Dignum, B. Dunin-Kęplicz and R. Verbrugge,
Creating collective intention through dialogue, *Logic Journal of the IGPL* 9 (2001), 145–158

B. Dunin-Kęplicz and R. Verbrugge,
Collective intentions, *Fundamenta Informaticae*, 51(3) (2002), 271–295

Forthcoming

B. Dunin-Kęplicz and R. Verbrugge,
Collective commitments in cooperative teams

B. Dunin-Kęplicz and R. Verbrugge,
A Tuning machine for collective commitments

B. Dunin-Kęplicz and R. Verbrugge,
Dynamics of collective attitudes during teamwork