From cognitive trust theories to computational trust

Jomi F. Hübner[1], Emiliano Lorini[2],
Laurent Vercouter[1], and Andreas Herzig[2]

[1]ENS Mines Saint Etienne, France
{hubner,vercouter}@emse.fr

[2]IRIT, Toulouse, France
{lorini,herzig}@irit.fr

Trust Workshop @ AAMAS 2009, Budapest

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

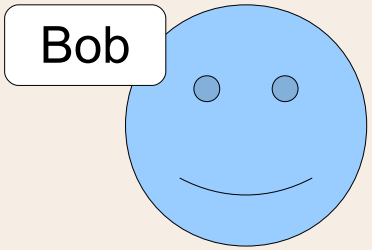
Outline

– Evaluate a cognitive definition of trust in ART scenario –

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

ART Testbed Scenario

Bob is asked to appraise a painting,
but he is not allowed to do it himself

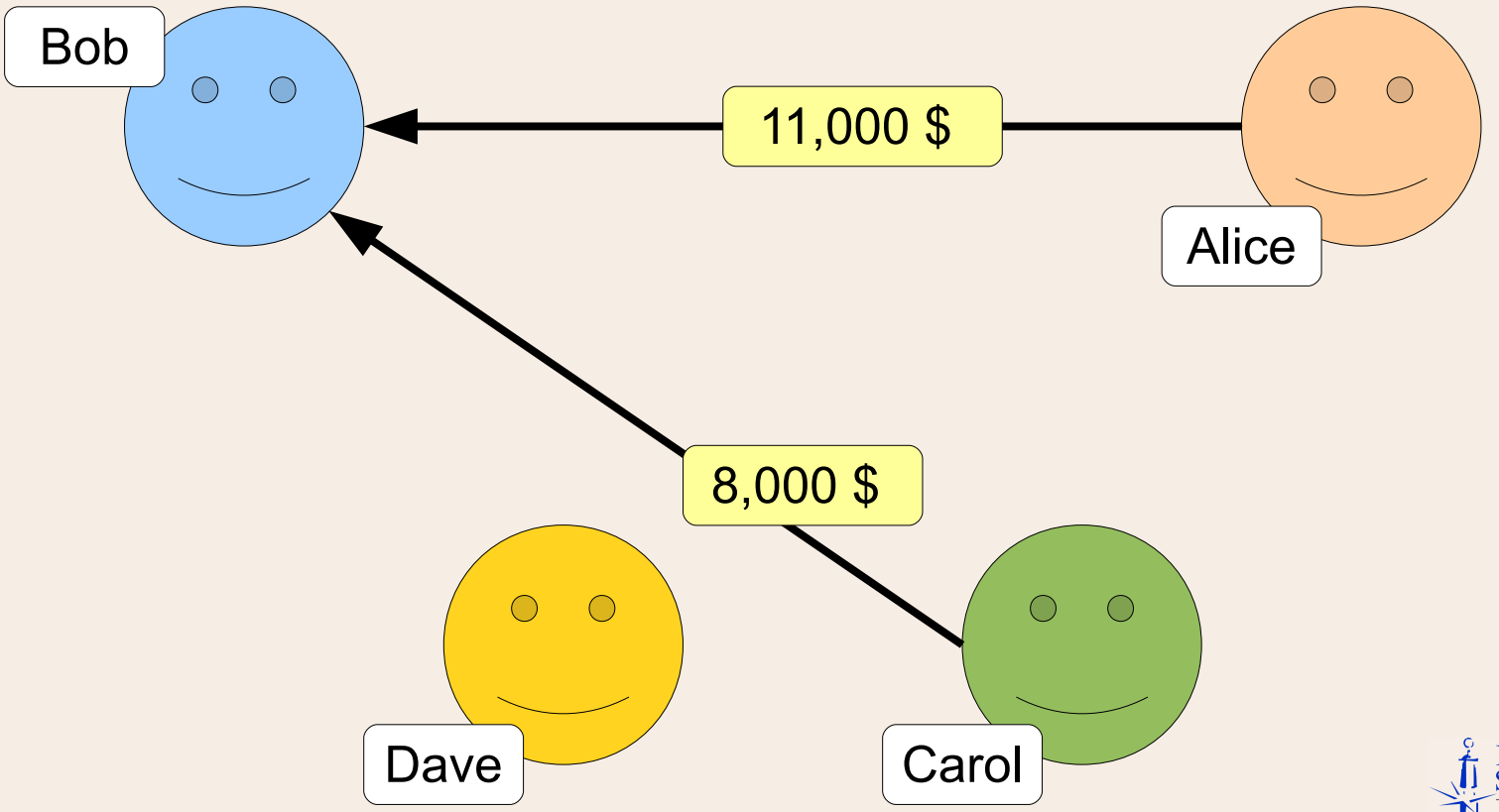


?

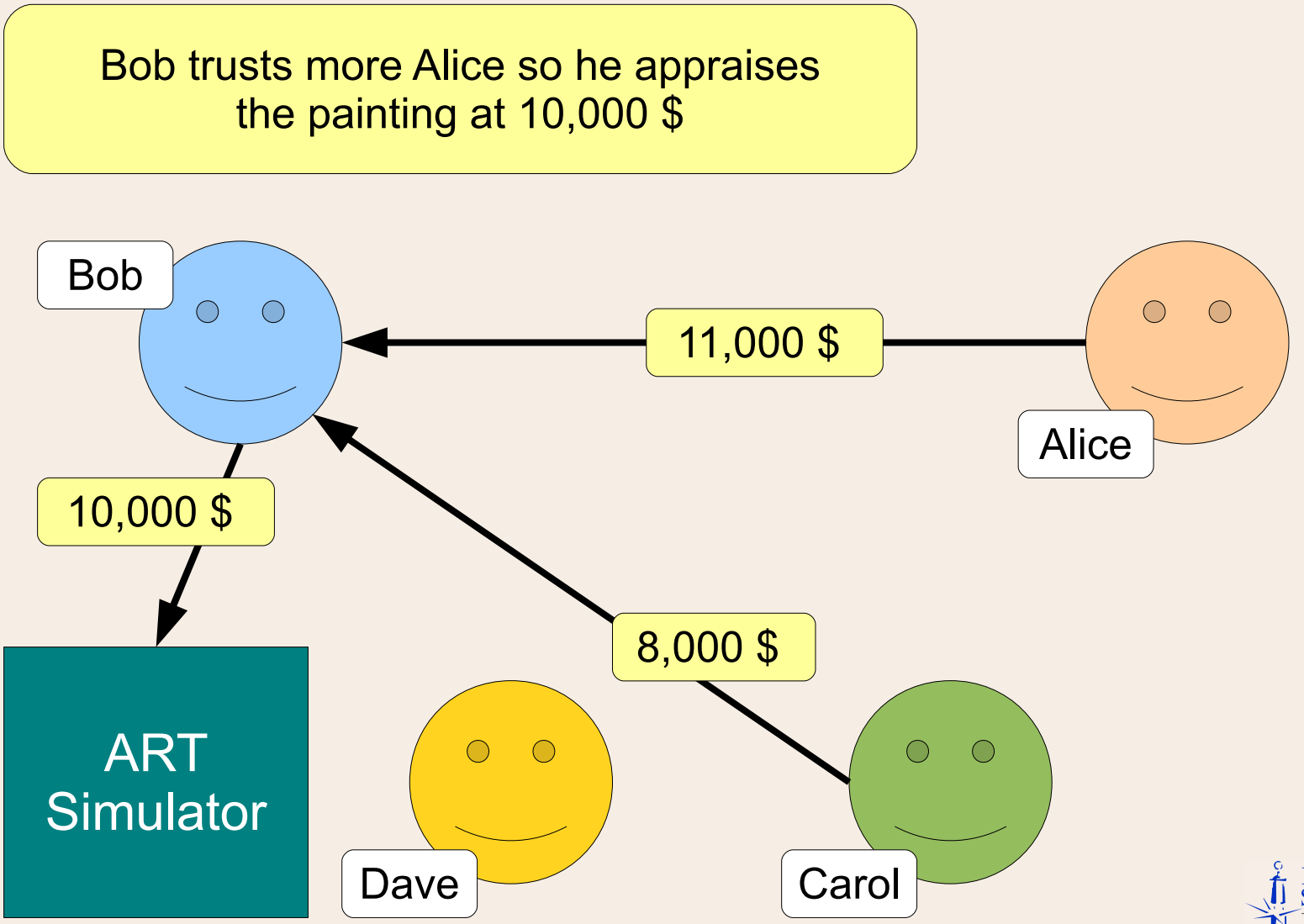
Ecole Nationale Supérieure des Mines SAINT-ETIENNE

ART Testbed Scenario

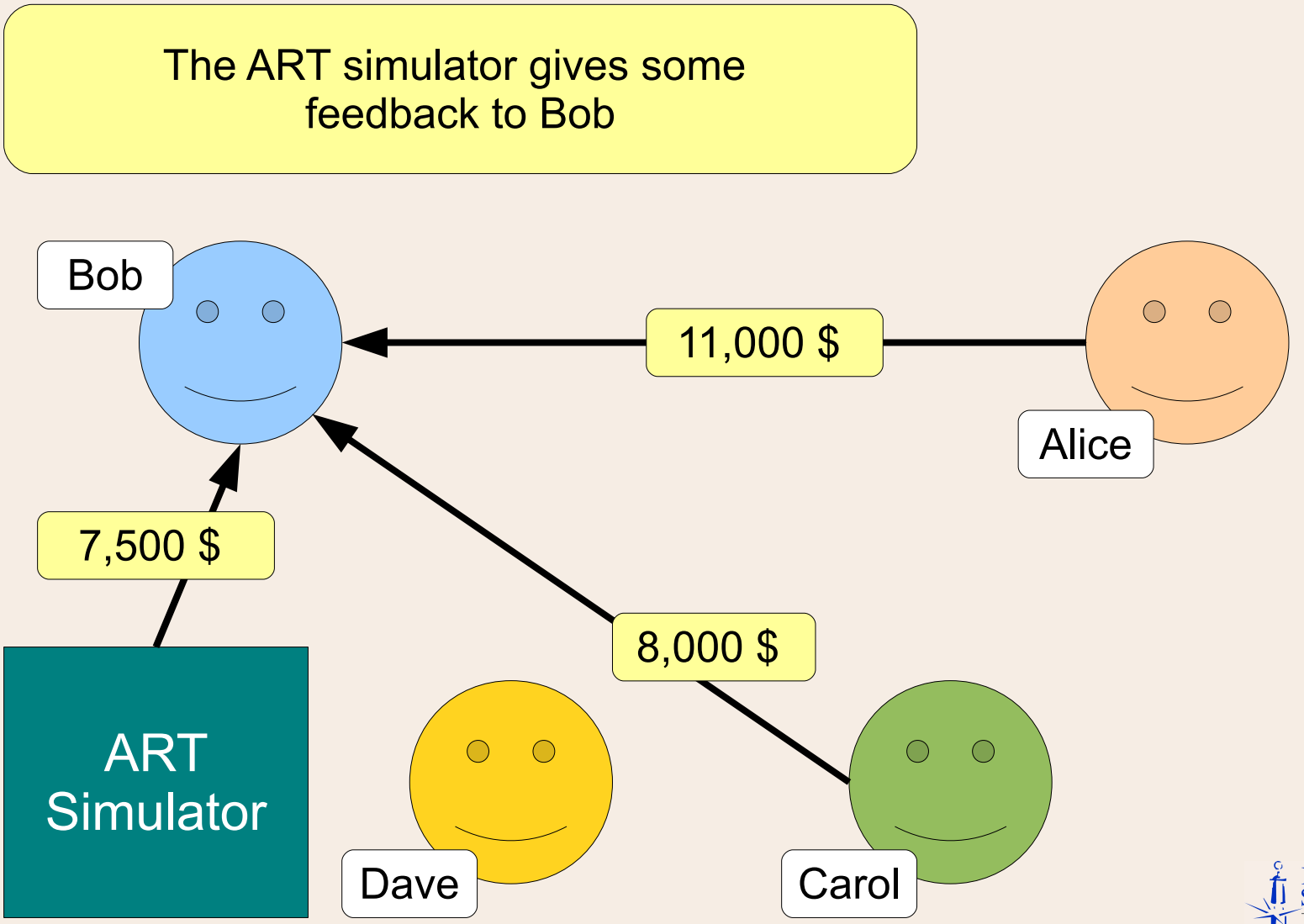
Bob will ask Alice and Carol for an opinion about the painting, he trusts them



ART Testbed Scenario

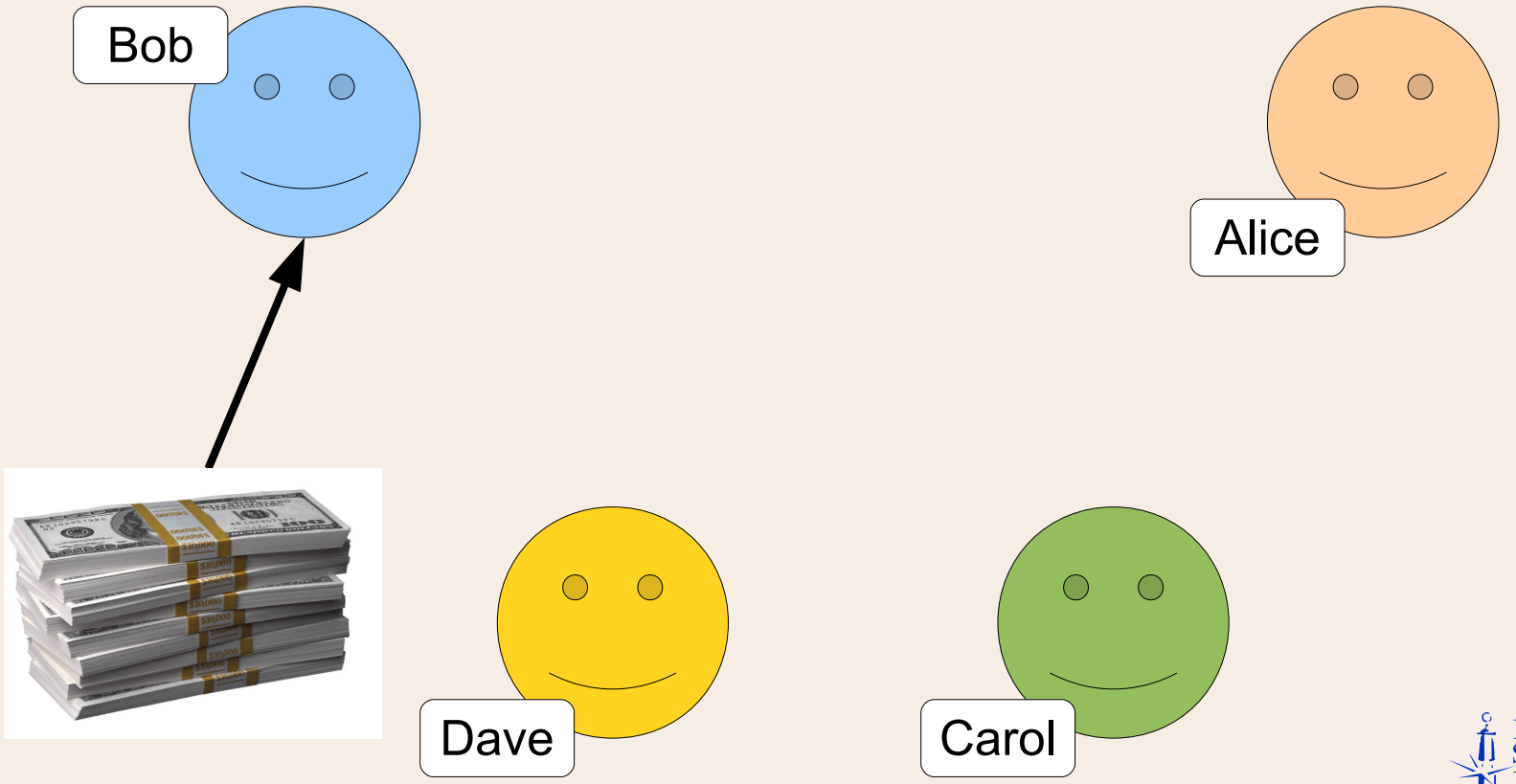


ART Testbed Scenario



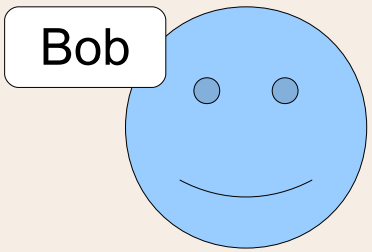
ART Testbed Scenario

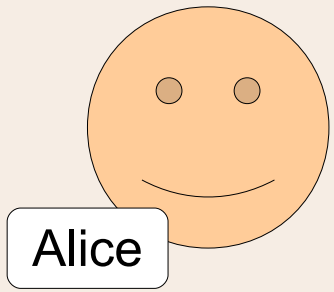
Bob is paid for its evaluation

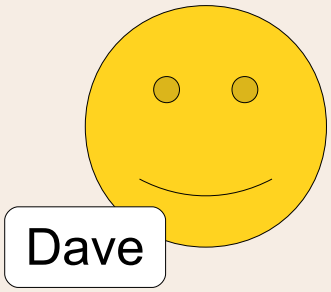


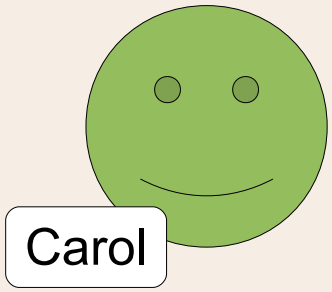
ART Testbed Scenario

Next Cycle start
The number of new clients for Bob depends on the quality of his evaluations in the previous cycle



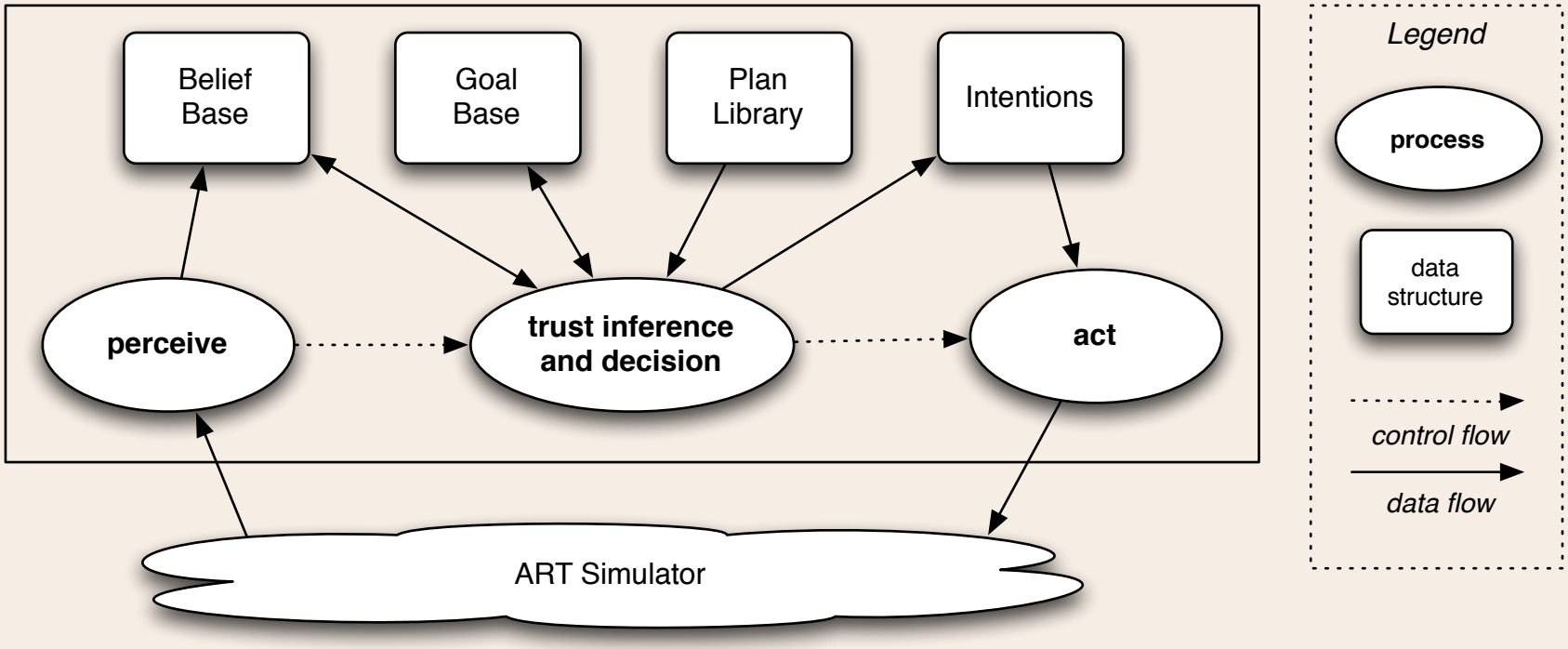






Ecole Nationale Supérieure des Mines SAINT-ETIENNE

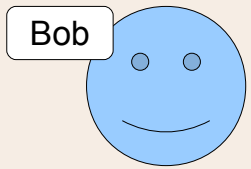
General Agent Architecture



Example of Perception

begin of cycle

Bob is asked to appraise a painting,
but he is not allowed to do it himself



?

painting(e, p, t)

- the painting p of era e is allocated to the agent at the current step t of the simulation

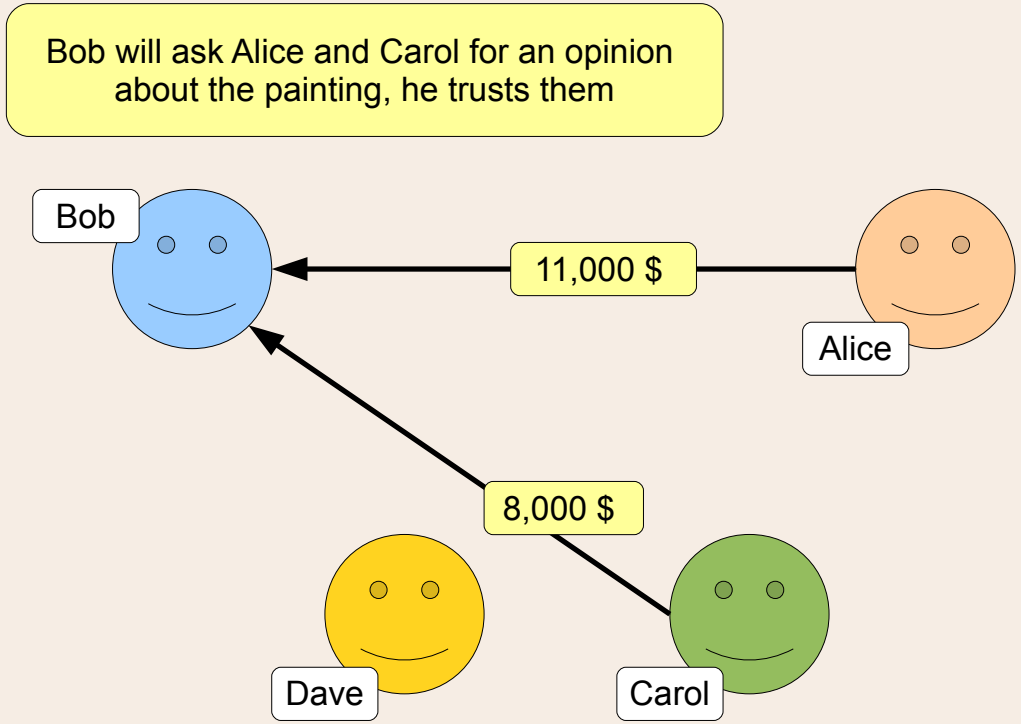
e.g. *painting*$(e1, p2345, 5)$

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Example of Action
ask opinion

$.send(j, ask, opinion(p))$

- ask agent j for an opinion about painting p



$.send(alice, ask, p)$
$.send(carol, ask, p)$

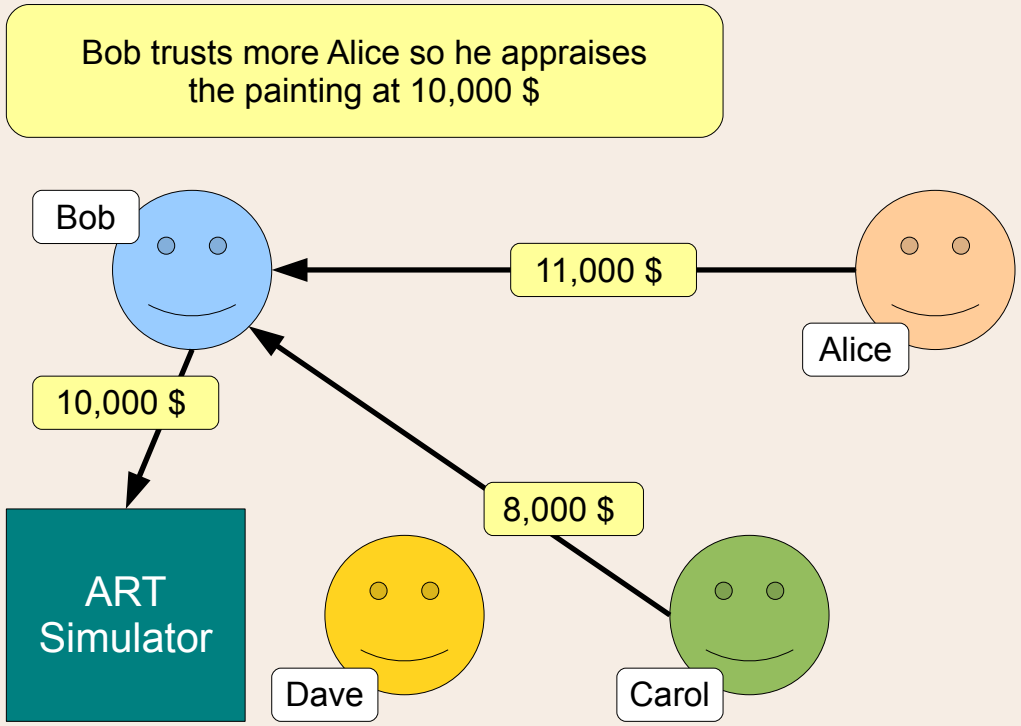
Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Example of Action

evaluate a painting

$opinion\_weight(j, p, x)$

- our agent evaluates the opinion of agent j with quality x



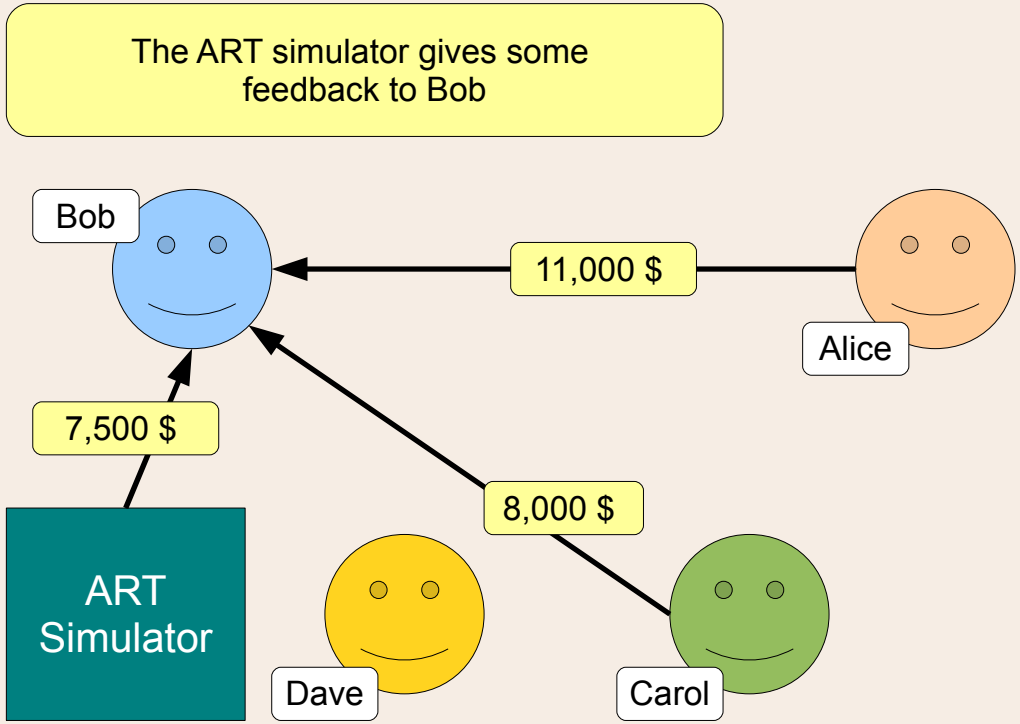
$opinion\_weight(alice, p, 0.9)$
$opinion\_weight(carol, p, 0.8)$

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Example of Perception
end of cycle

$opinion(j, e, v_g, v_r, t)$

- opinion produced by partner j for a painting of era e; the real value of the painting is v_r and the opinion provided by agent j is v_g



$opinion(alice, e1, 11000, 7500, 5)$
$opinion(carol, e1, 8000, 7500, 5)$

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

How to select partners to ask opinions?

Trust definition: Castelfranchi & Falcone

Trust has four components

- a truster i
- a trustee j
- an action α of j
- and a goal φ of i

"i trusts j to do α in order to achieve φ"

Trust's primitive ingredients

- i has the **goal** φ
- i believes that j is **capable** to do α
- i believes that j **intends** to do α
- i believes that j has the **power** to achieve φ by doing α

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Trust definition: Castelfranchi & Falcone

Trust has four components

- a truster i
- a trustee j
- an action α of j
- and a goal φ of i

"i trusts j to do α in order to achieve φ"

Trust's primitive ingredients

- i has the **goal** φ
- i believes that j is **capable** to do α
- i believes that j **intends** to do α
- i believes that j has the **power** to achieve φ by doing α

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Formalisation (ForTrust Project)

Definition (Occurrent Trust)

$$OccTrust(i,j,\alpha,\varphi) \stackrel{\text{def}}{=} Goal(i,\varphi) \wedge Believes(i, OccCap(j,\alpha)) \wedge Believes(i, OccIntends(j,\alpha)) \wedge Believes(i, OccPower(j,\alpha,\varphi))$$

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Formalisation (ForTrust Project)

Definition (Occurrent Trust)

$$\begin{aligned} OccTrust(i,j,\alpha,\varphi) \quad \stackrel{\text{def}}{=} \quad & Goal(i,\varphi) \wedge \\ & Believes(i, OccCap(j,\alpha)) \wedge \\ & Believes(i, OccIntends(j,\alpha)) \wedge \\ & Believes(i, OccPower(j,\alpha,\varphi)) \end{aligned}$$

Definition (Occurrent Trust — Dynamic Logic)

$$\begin{aligned} OccTrust(i,j,\alpha,\varphi) \quad = \quad & \mathtt{Pref}_i\, \mathtt{Eventually}\, \varphi \wedge \\ & \mathtt{Bel}_i\, \mathtt{Does}_{j:\alpha}\, \top \wedge \\ & \mathtt{Bel}_i\, \mathtt{After}_{j:\alpha}\, \varphi \end{aligned}$$

Formalisation (ForTrust Project)

Definition (Occurrent Trust — revised)

$$OccTrust(i,j,\alpha,\varphi) \stackrel{\text{def}}{=} Goal(i,\varphi) \wedge Believes(i, OccAct(j,\alpha)) \wedge Believes(i, OccPower(j,\alpha,\varphi))$$

Definition (Occurrent Trust — Dynamic Logic)

$$OccTrust(i,j,\alpha,\varphi) = \mathtt{Pref}_i\,\mathtt{Eventually}\,\varphi \wedge \mathtt{Bel}_i\,\mathtt{Does}_{j:\alpha}\,\top \wedge \mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\,\varphi$$

'Instantiation' of trust for ART testbed

- Actions: appraise a painting
  - All agents are capable to appraise, but with different expertise
- Goal: has good evaluations for the clients' painting
  - An agent cannot appraise its own painting, it have to ask opinions from others (which are also competitors)

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

'Instantiation' of trust for ART testbed

- Actions: appraise a painting
  - All agents are capable to appraise, but with different expertise
- Goal: has good evaluations for the clients' painting
  - An agent cannot appraise its own painting, it have to ask opinions from others (which are also competitors)

Definition (Trust in ART)

$$OccTrust(i, j, appraise(p), good\_eval(p)) \stackrel{\text{def}}{=} Goal(i, good\_eval(p)) \wedge Believes(i, OccAct(j, appraise(p))) \wedge Believes(i, OccPower(j, appraise(p), good\_eval(p)))$$

'Instantiation' of trust for ART testbed

- Actions: appraise a painting
  - All agents are capable to appraise, but with different expertise
- Goal: has good evaluations for the clients' painting
  - An agent cannot appraise its own painting, it have to ask opinions from others (which are also competitors)

Definition (Trust in ART — with strength)

$$OccTrust(i, j, appraise(p), good\_eval(p), min(x, y)) \stackrel{\text{def}}{=} Goal(i, good\_eval(p)) \wedge Believes(i, OccAct(j, appraise(p)), x) \wedge Believes(i, OccPower(j, appraise(p), good\_eval(p)), y)$$

Inference of ingredients I

$$
\begin{aligned}
&OccTrust(i, j, appraise(p), good\_eval(p), min(x, y)) \\
&\stackrel{\text{def}}{=} \quad Goal(i, good\_eval(p)) \wedge \\
&\qquad Believes(i, OccAct(j, appraise(p)), x) \wedge \\
&\qquad Believes(i, OccPower(j, appraise(p), evaluate(p)), y)
\end{aligned}
$$

The verify whether $Goal(i, good\_eval(p))$ holds the agent simply consult its belief base

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Inference of ingredients II

$$
\begin{aligned}
&OccTrust(i, j, appraise(p), good\_eval(p), min(x, y)) \\
&\stackrel{\text{def}}{=} \quad Goal(i, good\_eval(p)) \wedge \\
&\qquad Believes(i, OccAct(j, appraise(p)), x) \wedge \\
&\qquad Believes(i, OccPower(j, appraise(p), evaluate(p)), y)
\end{aligned}
$$

$$
\begin{aligned}
&Believes(i, OccAct(j, \alpha), x) \\
&\leftarrow \quad Believes(i, opinions\_count(j, a, n), 1) \wedge \\
&\qquad a > 0 \ \wedge \ x = \frac{n}{a} \ \wedge \ x > \epsilon
\end{aligned}
$$

a is the number of opinions requested to j, n is the number of answers

$\rightsquigarrow$ j does α if he did in the past

Inference of ingredients III

$$
\begin{aligned}
&OccTrust(i, j, appraise(p), good\_eval(p), min(x, y)) \\
&\overset{\text{def}}{=} \quad Goal(i, good\_eval(p)) \wedge \\
&\qquad Believes(i, OccAct(j, appraise(p)), x) \wedge \\
&\qquad Believes(i, OccPower(j, appraise(p), good\_eval(p)), y)
\end{aligned}
$$

$$
\begin{aligned}
&Believes(i, OccPower(j, appraise(p), good\_eval(p)), y) \\
&\leftarrow \quad Believes(i, sincere(j), 1) \wedge \\
&\qquad Believes(i, painting(e, p), 1) \wedge \\
&\qquad y = image_t(j, e) \ \wedge \ y > \delta
\end{aligned}
$$

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Inference of ingredients IV

$image_t : AGT \times ERA \rightarrow [0, 1]$

$$image_t(j, e) = \begin{cases} 0.5 & \text{if } t = 0 \\ image_{t\text{-}1}(j, e) & \text{if } O_t^{j,e} = \emptyset \\ \gamma\ r_t(j, e) + (1\text{-}\gamma) image_{t\text{-}1}(j, e) & \text{otherwise} \end{cases}$$

$$r_t(j, e) = \frac{1}{|O_t^{j,e}|} \sum_{(v_g, v_r) \in O_t^{j,e}} 1 - \frac{|v_g - v_r|}{v_r}$$

$O_t^{j,e}$ is the set of all opinions provided by agent j to our agent in paintings of era e and simulation step t

v_g is the value provided by j

v_r the real value of the painting

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Jason implementation I

```
// -- Find a partner for painting P
+!find_candidate(P)
   : agents(Ags) &       // consult all agents from belief base
     .findall(opt(C,Ag), // create a list of trustful candidates
          .member(J, Ags) &
          not partner(J, P) &
                                 // consult the trust of J
          trust(J, appraise(P), good_eval(P))[strength(C)],
          Candidates) &
      Candidates \== []          // to use this plan, there
                                 // must be a candidate
   <- .max(Candidates,opt(C,J)); // get the best candidate
      +partner(J, P).            // add the assignment (partner)

// plan that randomly select a sincere partner
+!find_candidate(P)
   : random_ag(J) & not partner(J, P) & sincere(J)
  <- +partner(J, P).
```

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Jason implementation II

```
// trust inference rule, e.g. Act=appraise(p1), Goal=good_eval(p1)
trust(J,Act,Goal)[strength(C)] :-
   .intend(Goal) &                          // I have the goal
   occ_act(J,Act)[strength(X)] &            // J is capable and intend
   occ_power(J,Act,Goal)[strength(Y)] & // J has the power
   C = math.min(X,Y).                       // computes the strength of the trust
   // the strength of beliefs are represented by annotations, enclosed by [ ]

// when a painting is allocated to me, to evaluate it is a goal
+painting(Era,P) <- !good_eval(P).

// capability and intention are based on % of responses to requests
occ_act(J,appraise(P))[strength(X)] :-
   opinions_count(J,Asked,Provided) & Asked > 0 & X = Provided/Asked & X > 0.9.

// power is based on image and sincerity
occ_power(J, appraise(P), _)[strength(Y)] :-
   sincere(J) & painting(Era,P) & image(J, Era, Y) & Y > 0.5.
   // the image function is implemented as a belief where the third term is
   // the value of the image of agent J
```

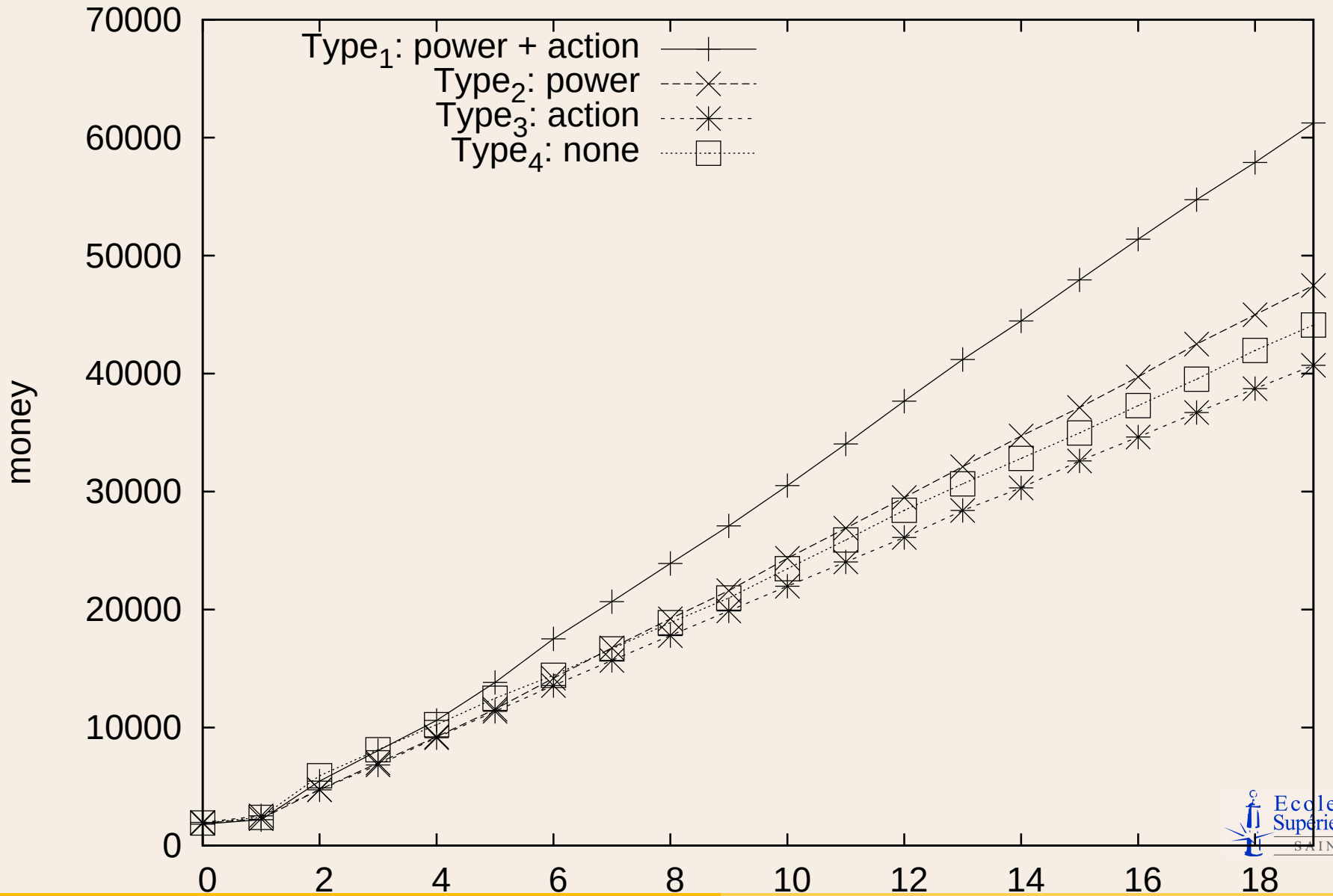
Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Jason implementation III

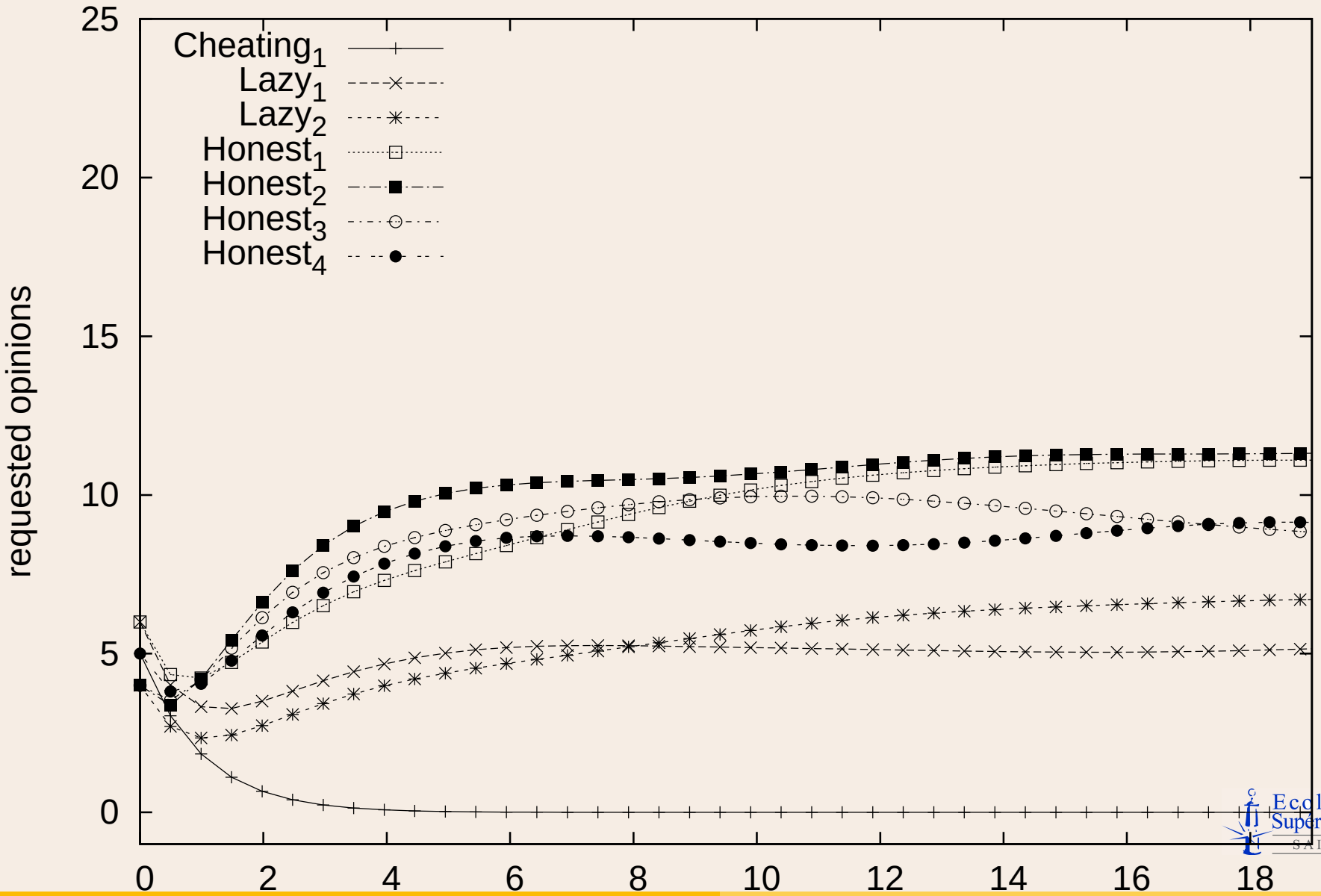
```
// whenever I receive an opinion from J
+opinion(J, Era, GivenValue, RealValue)
   <- Error = math.abs(RealValue - GivenValue) / RealValue;
      if (Error > 10) {  // huge errors means insincerity
         +~sincere(J)    // add a belief that J is not sincere
      };
      N = .count(opinion(J,Era,_,_)); // number of opinions
      R = (1-Error)/N;                // reward for the opinion
      ?image(J, Era, Img);            // consult current image
      NewImg = 0.5*Img + 0.5*R;       // compute new image
      -+image(J, Era, NewImg).        // update image belief
```

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

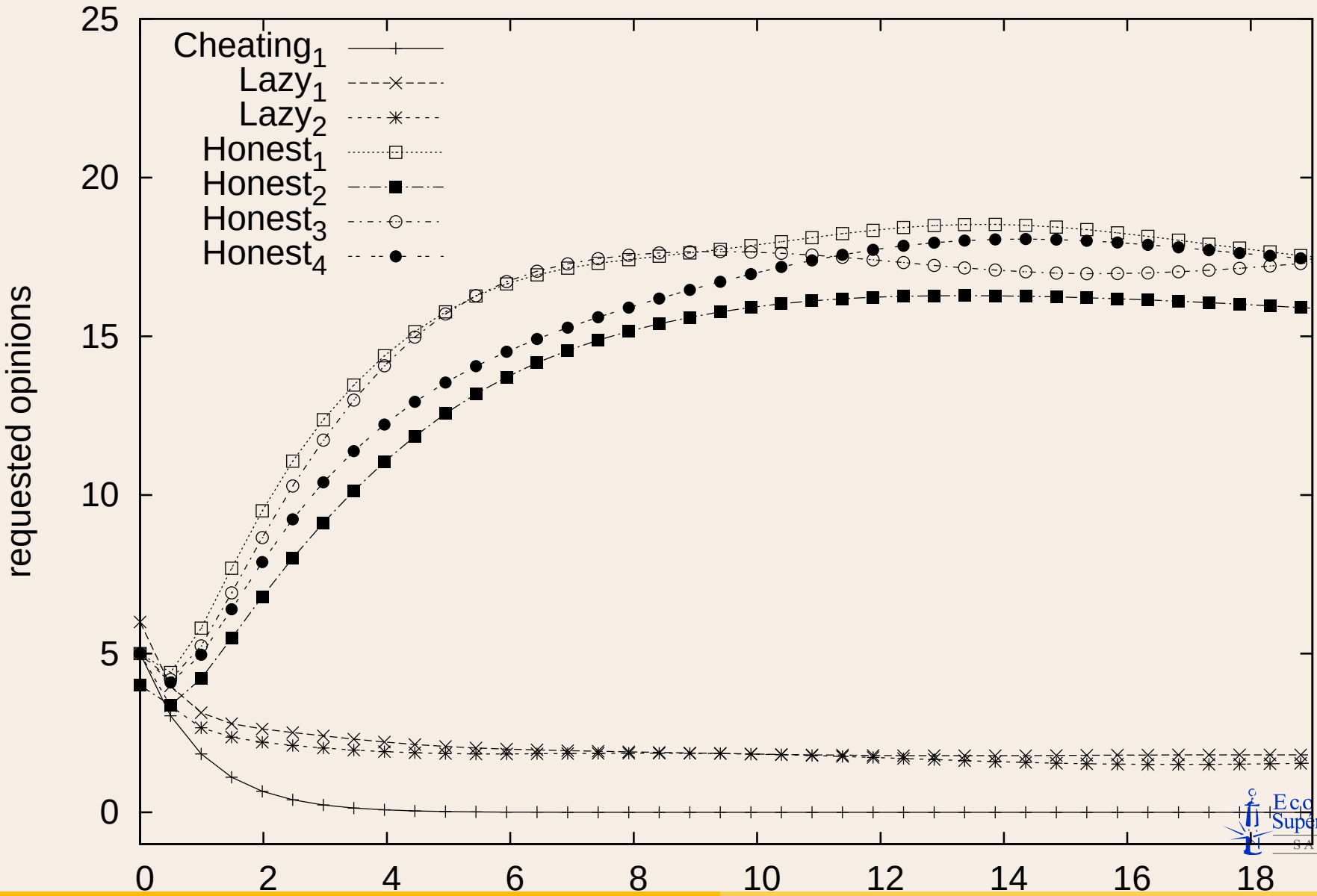
ForTrust vs ForTrust



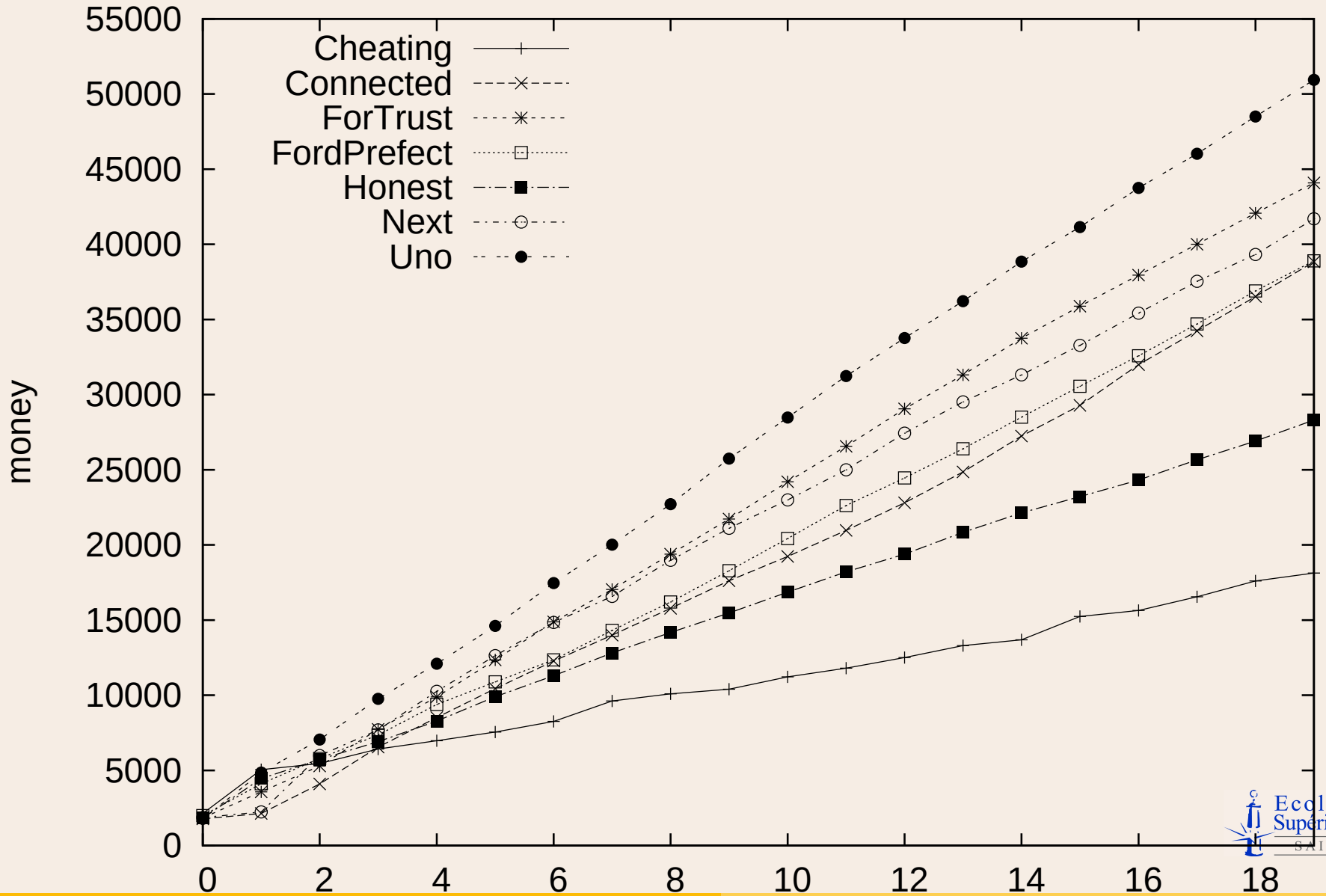
Partners of ForTrust-Power



Partners of ForTrust-Power-Action



ForTrust vs ART 2008 competitors



Summary

- From C&F to Logic
- From Logic to BDI architecture
- From BDI to Jason

- The proposal is
  - implementable
  - has good performance
  - all ingredients of the definition are relevant

- Importance of certainty (to rank partners)
- Adequacy to BDI languages

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Summary

- From C&F to Logic
- From Logic to BDI architecture
- From BDI to Jason

- The proposal is
  - implementable
  - has good performance
  - all ingredients of the definition are relevant

- Importance of certainty (to rank partners)
- Adequacy to BDI languages

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Summary

- From C&F to Logic
- From Logic to BDI architecture
- From BDI to Jason

- The proposal is
  - implementable
  - has good performance
  - all ingredients of the definition are relevant

- Importance of certainty (to rank partners)
- Adequacy to BDI languages

Ecole Nationale Supérieure des Mines SAINT-ETIENNE