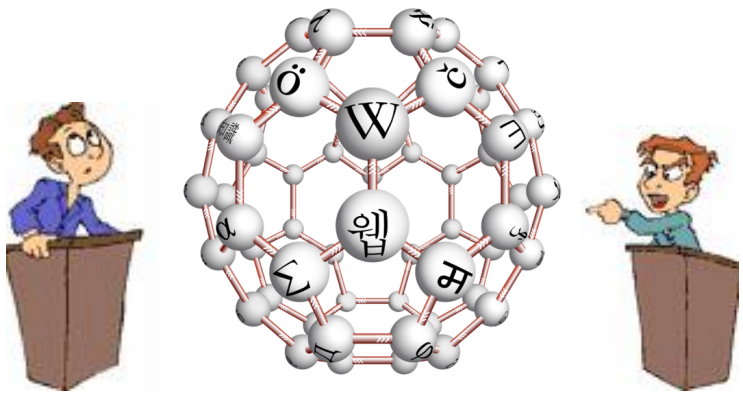
Justificatory and Explanatory Argumentation for Committing Agents

Adrian Groza

Department of Computer Science
Technical University of Cluj-Napoca, Romania
Adrian.Groza@cs.utcluj.ro



Outline

1. **Commitment-Based Justification&Explanation**
   - Justification-Explanation Complementarity
   - Enhancing Justification Logic with Explanation
   - Commitment Based Proof Terms
2. **Justificatory&Explanatory Patterns**
   - Justification Patterns
   - Explanation Patterns
3. **Argumentation Framework**
   - Arguing with Justification and Explanation
   - Argumentative Agents
4. **Trip Booking Scenario**

Justification-Explanation Complementarity

Humans tend to take decisions based both on knowledge and understanding.

Example (Judicial cases)

- circumstantial evidence for a delict needs to be complemented by a motive explaining the crime
- explanation itself is not enough without plausible evidence

In both situations the pleading is considered incomplete if either argumentation or explanation is missing.

A justification acts in a more normative framework, whilst an explanation works in a social context.
An explanation implies the existence of an audience which understands the claim and its explanans, whilst justification is in relation only with an objective world of true beliefs.

Justification and Explanatory Logic $\mathcal{JEL}$

- $\mathcal{JL}$ provides an evidence-based foundation for the logic of knowledge, according to which "F is known" is replaced by "F has an adequate justification"
- Instead of "X is known" (KX) consider $t : X$, that is, "X is known for the explicit reason t"
- Since justification and explanation have different functions we need two distinct operators to represent them: $t \triangleleft F$

Definition

The language of Justification and Explanation Logic $\mathcal{JEL}$ contains proof terms $t \in \mathcal{T}$ and formulas $F \in \mathcal{F}$

$$\begin{aligned} t: &= x \mid c \mid t \cdot t \mid t + t \mid !t \mid ?t \mid t \succ t \\ F: &= p \mid F \vee F \mid \neg F \mid t : F \mid t \triangleleft F \end{aligned}$$

Axioms of $\mathcal{JEL}$

A_0	classical propositional	axioms	
A_1	$F \rightarrow$	$(t : F \vee t \triangleleft F)$	(necessity)
A_2	$s : (F \rightarrow G) \rightarrow$	$(t : F \rightarrow (s \cdot t) : G)$	(j-application)
A_2'	$s \triangleleft (F \rightarrow G) \rightarrow$	$(t \triangleleft F \rightarrow (s \cdot t) \triangleleft G)$	(e-application)
A_2''	$s : (F \rightarrow G) \rightarrow$	$(t \triangleleft F \rightarrow (s \cdot t) \triangleleft G)$	(e-application)
A_2'''	$s \triangleleft (F \rightarrow G) \rightarrow$	$(t : F \rightarrow (s \cdot t) \triangleleft G)$	(e-application)
A_4	$t : F \rightarrow$	$!t : (t : F) \vee !t \triangleleft (t : F)$	(proof checker)
A_4'	$t \triangleleft F \rightarrow$	$!t \triangleleft (t \triangleleft F)$	(explanation checker)
A_5	$\neg t : F \rightarrow$	$?t : (\neg t : F) \vee ?t \triangleleft (\neg t : F)$	(negative proof checker)
A_5'	$\neg t \triangleleft F \rightarrow$	$?t \triangleleft (\neg t \triangleleft F)$	(negative explan. checker)

Commitment based justification in MAS

In a normative framework regulated only by social commitments, a justifier can be represented by such commitments: $C(x, y, Q, P)$

Expression	Informal Semantics
$C(a, b, p, q) :_i F$	C is probative evidence for F for agent i.
$\neg C(a, b, p, q) :_i F$	C is not probative justification for F for agent i.
$(\neg C(a, b, p, q)) :_i F$	The absence of C is a justification for F for agent i.
$C(a, b, p, q) :_i \neg F$	$C(a, b, p, q)$ is evidence for $\neg F$ for the agent i.

Feeding $\mathcal{JEL}$ machinery with commitments

Example (Distributed Application)

Assume the commitment $C(a, b, \_, F)$ is agent's b justification for F and the conditional commitment $C(b, c, F, G)$ represents a justification for agent c regarding the formula $F \rightarrow G$. According to axiom A_2, the application operator builds a justification for G of the form $C(b, c, F, G) \cdot C(a, b, \_, F) :_b G$.

$$C(b, c, F, G) :_c (F \rightarrow G) \rightarrow [C(a, b, \_, F) :_b F \rightarrow C(b, c, F, G) \cdot C(a, b, \_, F) :_b G]$$

Example (Positive Introspection)

Because agent a has promised agent b to deliver the *Item*, given by $C(a, b, \top, Item)$, this is the justification of agent a regarding the commitment $C(b, c, parts, pay)$ with agent c, in which agent b has to pay the components provided by agent c:

$$C(a, b, \top, Item) :_a [C(b, c, parts, pay) :_a Item]$$

.

Outline

1 Commitment-Based Justification&Explanation
- Justification-Explanation Complementarity
- Enhancing Justification Logic with Explanation
- Commitment Based Proof Terms

2 Justificatory&Explanatory Patterns
- Justification Patterns
- Explanation Patterns

3 Argumentation Framework
- Arguing with Justification and Explanation
- Argumentative Agents

4 Trip Booking Scenario

Gratuitous promise

Justification from gratuitous promise $\div$ *GP*

Intent:	*Stressing out that no one has forced the debtor to commit.*
Context:	*The creditor or other agent perform actions based on the promise.*
Pattern:	$C(a, b, \top, P) :_c F$
Variants:	*b=c, a=c, F=P*
Example:	*Grandfather promises his nephew to pay for a trip. Based on it, the father buys a new bag for his son.*
$\mathcal{JEL}$:	$C(grandfather, nephew, \top, PayTrip) :_{father} NewBag$

- CQ_1: If $c \neq b$, which is the relationship between the agent c and b?
- CQ_2: If $F \neq P$, which is the link between the formulas F and P?
- CQ_3: Is there a stronger commitment which does not justify F?

Requests

Justification from request $\doteqdot R$

Intent :	*Justifying actions based on the directive conveyed by a normative empowered agent.*
Context :	*The creditor or other agent act on request from other agent.*
Pattern :	$C(a, b, P, \top) :_c F$
Variants :	*b=c, a=c, F=P,* $\neg P$
Example :	*During driving lessons, the instructor requests the student to stop the car. Consequently, the next student's action is to signal right.*
$\mathcal{JEL}$:	$C(instructor, student, StopCar, \top) :_{student} SignalRight$
CQ_4 :	*Is the request legitimate?*

If the requested act is negated, the commitment represents a *taboo* or an interdiction.

Higher order patterns

1. unilateral contract *UL* - by composing a gratuitous promise with a fact Q treated as a condition:
   $C(a, b, \top, P) \circ_q Q = C(a, b, Q, P)$ or
   $C(a, b, Q, \top) \circ_p P = C(a, b, Q, P)$
2. bilateral contract *BC* - by composing an UC with a gratuitous promise

   $$C(a, b, C(b, a, \top, pay), deliver) = C(a, b, Q, P) \circ_q C(b, a, \top, Q_1)$$

3. promise to commit - composition of two gratuitous promises, applied on the fourth term of the commitment:

   $$C(a, b, \top, C(a, c, \top, deliver)) = C(a, b, \top, \_) \circ_p C(a, c, \top, pay)$$

Public goals

Explanation patterns are not rooted in an objective normative frame, having a subjective component. Explanatory schemes can be viewed as providing subjective reasons, with a more flexible relation between explainers and what is explained.

Explanation from cognitive_consistency $\doteqdot$ *CC*

Intent :	*Explaining actions based on the goals that an agent is following.*
Context :	*The debtor commits itself to achieve a particular sentence.*
Pattern :	$C(a, a, Q, P) \triangleleft_c F$
Variant :	*a=c,* $Q = \top$
Example :	*Tom cannot join the party because he wants to learn for the exam.*
$\mathcal{JEL}$:	$C(tom, tom, \top, Learn) \triangleleft_{jim} \neg JoinParty$
CQ_4:	*Is the debtor aware of his commitments?*

Explanation from preference

Explanation from preferred_commitment $\doteqdot$ *PC*

Intent :	*Explaining choice between two commitments.*
Context :	*The debtor is committed with different strengths to creditors.*
Pattern :	$C(a, b, P, Q) \succ C(a, c, P', Q') \lhd_d F$
Variants :	*a=d, b=d, c=d*
Example :	*The agent "a" promised his boss to attend a late meeting. He also promised his wife to take the child from school if he has time. Aware of the constraint, the wife decides to go to the school herself.*
$\mathcal{JEL}$:	$C(a, boss, \top, Meeting) \succ C(a, wife, Time, TakeChild) \lhd_{wife} GoSchool$
CQ_4 :	*Is it not possible to achieve both commitments?*
CQ_5 :	*Is the preference relation explained by the cognitive consistency property of the agent d?*

Outline

1 Commitment-Based Justification&Explanation
- Justification-Explanation Complementarity
- Enhancing Justification Logic with Explanation
- Commitment Based Proof Terms

2 Justificatory&Explanatory Patterns
- Justification Patterns
- Explanation Patterns

3 Argumentation Framework
- Arguing with Justification and Explanation
- Argumentative Agents

4 Trip Booking Scenario

Definition (Justificatory and explanatory arguments)

An argument is a pair $\langle t, F \rangle$ where t is the chain of justifiers or explanans and F the conclusion such that t justifies or explains F. A justificatory argument is supported by justifiers only. An explicatory argument contains at least one supporting explanandum.

Definition (Conflict among arguments)

In the case of an undercutting argument, its conclusion attacks one justifier or explanation in the support of another argument. In the case of rebuttals, the justified formulas contradict each other directly.

Definition (Request for justification)

Agent a requests agent b to provide its justification why t is agent's b justification for F: $C(a, b, !t :_b t :_i F, \top)$.

Definition (Request for explanation)

Agent a requests agent b to explain why t is the justification of i for F: $C(a, b, !t \lhd_j t :_i F, \top)$. Similarly, agent a requests agent b to further explain why t is the explanation of agent i for F: $C(a, b, !t \lhd_j t \lhd_i F, \top)$.

Strength of justification

The stronger the commitment, the stronger the justification.
The commitment $C(a, b, q, p)$ is stronger than $C(a, b, q', p')$ if the debtor promises more and requests less.

Example (Strength of justification)

"The supplier s commits to deliver more items faster to the retailer r" is stronger than "The supplier commits to deliver the items earlier if he receives the payment earlier":

$$C(s, r, \top, \mathit{MoreItems} \wedge \mathit{FasterDeliv}) \succ C(s, r, \mathit{PayEarlier}, \mathit{FasterDeliv})$$

Example (Strength of justification)

The request to pay either by credit card or by wire transfer is stronger than the wired transfer option, formalised as:
$C(\mathit{bank}, a, \mathit{card} \vee \mathit{wired}, \top) \succ C(\mathit{bank}, a, \mathit{wired}, \top)$.

Preference over commitments

A fact p is stronger than the commitment created to bring about that p: $p \succcurlyeq C(\_, \_, \_, p)$.

Example

"I commit to deliver the item after you confirm the order" is stronger then "I commit to deliver the item if you have promised me to confirm the order": $C(me, you, Confirm, Deliver) \succcurlyeq C(me, you, C(me, you, \top, Confirm), Deliver)$.

The creditor prefers stronger commitments, whilst the debtor prefers weaker commitments when justifying the formula F.
If $C(a, b, P, Q) \succcurlyeq C(a, b, P', Q')$ then
$C(a, b, P, Q) \succ C(a, b, P', Q') \rhd_a F$ and
$C(a, b, P', Q') \succ C(a, b, P, Q) \rhd_b F$
In case the justifications have the same strength, $\mathcal{JEL}$ forces the agents to provide further justifications on request.

Types of agents

Exploiting $\mathcal{JEL}$ axioms for defining agents

- The less demanding agent requires only explanations in order to accept a formula: $t \triangleleft_i F \rightarrow F$.
- A *rigorous* agent accepts only normative justified formulas: $t :_i F \rightarrow F$.
- The most demanding one requests both explanation and justification before accepting a sentence: $t \triangleleft_i F \wedge s :_i F \rightarrow F$.
- A *caution* agent, which accepts a formula F if it has a valid justification and no valid explanation supporting the opposite conclusion $\neg F$:

$$t :_i F \wedge (\neg s) \triangleleft_i \neg F \rightarrow F$$

Exploiting existing theories of justification (foundationalism, infinitism, internalism, externalism) and explanation (causal, teleological)

- A *foundationalist* agent, the existence of a basic justificatory pattern would be enough for accepting the supported formula.
- An *n-type credulous* agent accepts a formula if its justification chain has at least length n.
- An *internalist* agent should be able to justify a sentence only through its own commitments

Standard of Proofs

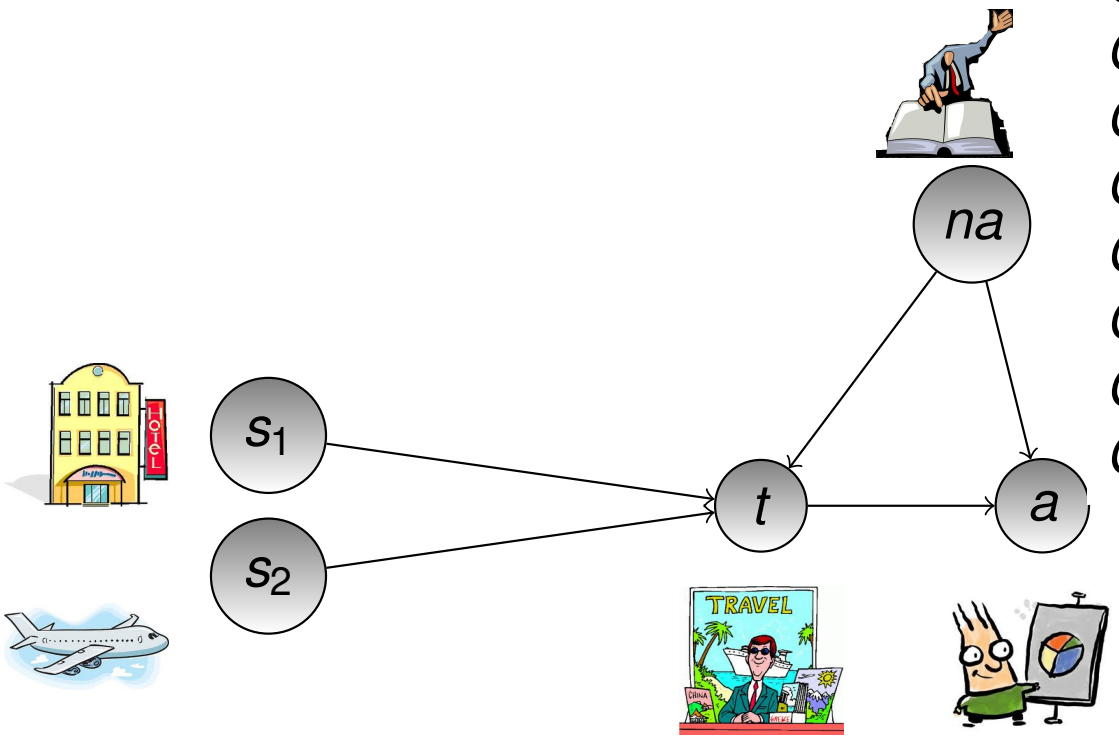
Agents can convey or accept as valid justifiers only commitments meeting a strength threshold.

Pattern/Standard	**Justifier**	**Meaning**
$GP + (R \circ_q GP)$ (scintilla of justification)	$C(a, b, \top, deliver)+$ $C(a, b, C(b, a, \top, pay), \top) :_a Pay$	a commits to deliver the item & requests b to commit to pay it.
$GP + R$ (reasonable justification)	$C(a, b, \top, deliver)+$ $C(a, b, pay, \top) :_a Pay$	a commits to deliver the item and requests b to pay for it.
$UC \circ_q GP$ (preponderence of justification)	$C(a, b, C(b, a, \top, pay), deliver) :_a Pay$	a commits to deliver the item if b commits a to pay for it.
UC (convincing justification)	$C(a, b, pay, deliver) :_a Pay$	a commits to deliver the item in case b pays for it.
$UC \circ_p GP$ (behind any reasonable doubt)	$C(a, b, pay, C(a, b, \top, deliver)) :_a Pay$	a will commit to deliver the item if b pays.

Outline

1. Commitment-Based Justification&Explanation
   - Justification-Explanation Complementarity
   - Enhancing Justification Logic with Explanation
   - Commitment Based Proof Terms
2. Justificatory&Explanatory Patterns
   - Justification Patterns
   - Explanation Patterns
3. Argumentation Framework
   - Arguing with Justification and Explanation
   - Argumentative Agents
4. Trip Booking Scenario

Commitment store in the running scenario.



Commitment Store:

$C(t, a, pay_3, trip)$

$C(t, t, flight \wedge acc, trip)$

$C(s_2, t, pay_1, acc)$

$C(s_1, t, C(t, s_1, \top, pay_2), flight)$

$C(na, t, \neg trip \wedge pay_3, C(t, \_, \top,$

$C(na, \_, \neg EUcitizen, visa)$

$C(na, \_, swiss, \neg visa)$

Illustrative scenario

internalist, 1-type credulous
convincing justification

externalist, cautious
preponderance of evidence

The agent a commences the dialogue by requesting justification j guaranteeing the trip in case he pays, given by the R pattern:

$$C(a, t, j :_t \mathit{Trip}, \top)$$

Being an internalist agent, the tourism company t can guarantee the trip based only on its contractual clauses with the suppliers s_1 and s_2:

$$[C(s_2, t, \mathit{pay}_1, \mathit{acc}) + C(s_1, t, C(t, s_1, \top, \mathit{pay}_2), \mathit{flight})] :_t \mathit{Trip}$$

We note the compound justifier above with J, so that $J :_t \mathit{Trip}$. The first term in J is an UC, whilst the second commitment represents a BC. Both terms satisfying the proof standard of agent a, the compound justifier J will meet the justification standard of agent a.

Illustrative scenario

internalist, 1-type credulous convincing justification

externalist, cautious preponderance of evidence

Being a cautious agent, the client should check if there are explanations supporting the opposite conclusion $\neg Trip$. Firstly, it requests explanans why J offers enough justification for the formula $Trip$:

$$C(a, t, !J \lhd_t J :_t Trip, \top)$$

The tourism agency explains based on its business practice that plane tickets and accommodation are enough to provide the requested trip:

$$C(t, t, acc \wedge flight, trip) \lhd_t J :_t Trip$$

Consider that the agent a is aware of the regulation requesting visa for non EU citizens, formalized as $C(na, \_, \neg EUcitizen, visa)$. This is the main concern of the agent a for rebutting the $Trip$ formula:

$$C(na, \_, \neg EUcitizen, visa) :_a J :_a \neg Trip$$

We note the above justification chain with J' such that $J' :_a \neg Trip$. The tourism agency may request explanations to agent a regarding his concern:

$$C(t, a, !J' \lhd_a J' :_a \neg Trip, \top)$$

Illustrative scenario

internalist, 1-type credulous
convincing justification

externalist, cautious
preponderance of evidence

The agent a can provide the explanation that he is not an EU citizen. This helps agent t to figure out the case, but being a 1-type credulous agent it needs a justification to accept the formula $\neg EUcitizen$:

$$C(t, a, J'' :_a \neg EUcitizen, \top)$$

Agent a may provide an ID card or passport acting as a justification: $swissPassport :_a \neg EUcitizen$. A fact being stronger than a commitment, it satisfies the convincing justification standard of the t agent.
Swiss citizens do not need a visa for travelling in Europe, the commitment $C(na, \_, swiss, \neg visa)$ undercuts the $a's$ justifier supporting $\neg Trip$:

$$C(na, \_, swiss, \neg visa) :_t \neg J' :_a \neg Trip$$

We note the above justifier with J''' such that $J''' :_t \neg J' :_a \neg Trip$.

Illustrative scenario

internalist, 1-type credulous
convincing justification

externalist, cautious
preponderance of evidence

Being cautious, the agent *a* needs a supplementary explanation, requested with:

$$C(a, t, !J''' \lhd_t J''' :_t \neg J' :_a \neg Trip, \top)$$

The EU preferring to encourage travelling across Europe from safe countries, rather then imposing unnecessary security constraints, explains why a visa is not required:

$$[C(na, na, \top, encourageTravel) \succ C(na, \_, \neg EUcitizen, visa)] \lhd_t J''' :_t \neg J' :_a \neg Trip$$

Having both a justification and an explanation, the agent *a* accepts that his visa concerns are defeated.
Consequently, given the accepted valid justification *J* for the trip, and no valid explanation for $\neg Trip$, the agent accepts the *Trip* formula.

Contributions

- setting some basis of exploiting, in computational models of arguments, the differences between justification and explanation, as already stressed out in the philosophy of science.
- the $\mathcal{JEL}$ is developed to cover both justifiers and explanans, in the line of using logic in argumentation
- formalisation of justificatory and explanatory commitment-based patterns
- using together $\mathcal{JEL}$ with commitments provides opportunities for defining several types of argumentative agents.

Thank you!