

Logical Judges Challenge Human Judges on the Strange Case of B.C.–Valjean

Viviana Mascardi

University of Genova, DIBRIS, Italy
viviana.mascardi@unige.it

Domenico Pellegrini

Ministry of Justice, Tribunale di Genova, Italy
domenico.pellegrini@giustizia.it

On May 12th, 2020, during the course entitled *Artificial Intelligence and Jurisdiction Practice* organized by the Italian School of Magistracy, more than 70 magistrates followed our demonstration of a Prolog logical judge reasoning on an armed robbery case. Although the implemented logical judge is just an exercise of knowledge representation and simple deductive reasoning, a practical demonstration of an automated reasoning tool to such a large audience of potential end-users represents a first and unique attempt in Italy and, to the best of our knowledge, in the international panorama. In this paper we present the case addressed by the logical judge – a real case already addressed by a human judge in 2015 – and the feedback on the demonstration collected from the attendees.

1 Introduction

The connections between logic programming and law have been studied for a long time. In 1975, Meldman discussed his PhD Thesis entitled “A preliminary study in computer-aided legal analysis” [12] where he modelled legal facts in a Lisp-like language and used instantiation (recalling unification) and syllogism (recalling resolution) to perform a simple kind of legal analysis inspired by Prosser’s Law of Torts [13]. At that time Prolog was just born, but its applications to legal reasoning were not long in coming. One of the first attempts was made by Hustler [9] who implemented a prototype of a legal consultant in Prolog, again inspired by Prosser’s work. A few years later, Kowalski, Sergot et al. succeeded in running a significant portion of the 1981 British Nationality Act, implemented in Prolog on a small micro computer [15]. In the same years, Prolog became very popular for implementing expert systems for the legal domain [3, 19].

From those early attempts, much progress has been made: research on deontic and defeasible reasoning [1, 5], ontological reasoning [7], and argumentation [8, 18] is extremely lively and helps disclosing the many connections between logic programming (and, more in general, computational logic and automated reasoning) and legal reasoning. The application of automated reasoning to digital forensics is another promising research direction [6] whose potential is witnessed by the ongoing “Digital Forensics: Evidence Analysis via Intelligent Systems and Practices” (DigForASP) COST Action¹. DigForASP exploits computational logic to reason on crimes evidences to reconstruct possible scenarios related to the crime, even when knowledge is fragmented and incomplete.

Despite this long and successful history, the potential and limitations of automated reasoning are still obscure to their end-users, namely judges, magistrates, lawyers and prosecutors. These professionals are overwhelmed by news about robotic judges but are not always fully aware of the techniques behind these robotic surrogates; evaluating their pros and cons in an informed and objective way is often out of their reach.

¹CA17124, <https://digforasp.uca.es>, funded for four years starting from 09/2018 by the European Cooperation in Science and Technology (COST, www.cost.eu).

At the time being, all the “robotic judges” used in trials employ machine learning, most often deep learning, and many of them became famous for their biased decisions. The *State v Loomis 881 N.W.2d 749 (Wis. 2016)* case is one among the most well known examples: the Wisconsin Supreme Court upheld a lower court’s sentencing decision informed by a COMPAS risk assessment report and rejected the defendant’s appeal on the grounds of the right to due process. COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a case management and decision support tool developed and owned by Equivant². It is opaque from two points of view: legal, because it is a commercial software whose source code cannot be inspected, and technical, because it employs machine learning techniques which, in many cases including their more recent and successful “deep” evolution, are black boxes [10]. Using machine learning for boosting predictive justice is becoming a very lively research field, although in many cases the developed applications are academic prototypes, not yet used in real trials. Applications range from predicting decisions of the European Court of Human Rights [11] to predicting recidivism of many different crimes [4], to risk assessment in criminal justice [2].

On the one hand, many non technical papers foresee the rise of robotic judges suggesting that they might substitute human judges in most of their activities. On the other, many scientists warn about opaque predictive models also from a technical point of view, besides an ethical one [16], and advocate the adoption of interpretable models instead [14]. In between, human judges, whose computer literacy is often a basic one, are more and more confused.

To address this pressing need of clarity, on the 12th and 13th of May 2020 the Italian School of Magistracy (Scuola Superiore della Magistratura, SSM) offered a course entitled *Artificial Intelligence and Jurisdiction Practice*. Within that course, we were in charge for the working group on *Criminal Law* carried out during the afternoon of May 12th. The working group was run in parallel with the *Civil Law* group and involved more than 70 attendees, one half of the total number of attendees of the course. The design of the working group activities required some initial effort due to the mismatches between the vocabulary of the two authors – a magistrate and a computer scientist – with a completely different background. Once we aligned our shared terminology, the set up of the activities and the preparation of the teaching material run smoothly and many connections between the Italian law and computational logics were discovered, from the adoption of *modus ponens* as a well known oratory technique [17] to closed world assumption³.

To allow the attendees to have an idea of how computational logic might serve their needs, having less than two hours at our disposal, one week before the course took place we sent them an exercise based on a real, published case of armed robbery, properly obfuscated to avoid that they could recognize it. They had all the details of the case and they were asked to answer some questions based on the availability of evidences and on the reliability of witnesses. In the meanwhile, we translated the robbery facts and evidences into Prolog and we implemented the logical judge. During the working group, after an introductory presentation to artificial intelligence and logic programming, we run a demo to show the conclusions reached by the logical judge based on the available evidences, and we entered a discussion on whether, and how, the logical judge could reach the same conclusions as the human ones.

Although our logical judge is no more than a simple exercise of knowledge representation in Prolog, showing what automated reasoning in the legal domain could achieve via a practical demonstration to more than 70 magistrates represents a first and unique attempt in Italy and, to the best of our knowledge, in the international panorama.

²<https://www.equivant.com/>, last accessed June 2020.

³The article 530, second paragraph, of the Italian Code of Criminal Procedure states that “the judge acquits the defendant also in case *there is no evidence of the crime, or the evidence is not sufficient, or it is contradictory*”, which closely resembles Prolog negation as failure and closed world assumption.

2 The Human Judge and the Case of B.C.

The armed robbery case we took inspiration from is the case of B.C.⁴ To avoid disclosing the case, we named B.C. *Valjean* and we used other names from *Les Misérables* by Victor Hugo for other actors. We also added some further evidences and conditions to make the reasoning more involved⁵ The revised B.C. case is the following. A criminal wearing a red jacket and a full-face motorcycle helmet enters the ABC supermarket wielding a gun, together with his partner in crime. He threatens Enjolras with the gun, asks for the money, and hits him. The two criminals try to get the money but, due to the prompt and unexpected reaction of Enjolras, after a few minutes run away on a scooter. During the trial, the following evidences emerge:

- **E1**: Fantine, a highly reliable witness, declares that the two criminals left the supermarket on board of a scooter with plate 12345 at 15.00, more or less. The scooter – whose theft had been reported a few days before – was found later in the afternoon.
- **E2**: on the scooter’s rearview mirror, the scientific police finds a fingerprint highly compliant with Valjean’s one.
- **E3**: Fantine also asserts that, before leaving the supermarket, the criminal with the red jacket said to his partner “Jamunindi, jamunindi!”. This sentence means “Let’s go” in the dialect from Reggio Calabria, and Valjean was born in Reggio Calabria.
- **E4**: Thenardier asserts that he saw Valjean at 15.05 riding a scooter with plate 12345, in a road very close to the ABC supermarket, together with another man.
- **E5**: in a sound track extracted from a mobile phone accidentally retrieved in the supermarket proximity, dating back 14.55 of the robbery day, a voice that turned out to be Valjean’s one can be heard.
- W.r.t. **E4**, the defense lawyer presents evidences that Thenardier is an unreliable witness.

The exercise we proposed to the attendees is the following: based on the declarations of the witnesses, can we demonstrate that the criminal in red jacket is Valjean...

- **Q1** ...using evidences **E1, E2, E3**?
- **Q2** ...using evidences **E1, E2, E3, E4**, without using the evidences from the defense lawyer on Thenardier’s unreliability?
- **Q3** ...using evidences **E1, E2, E3, E4** and evidences from the defense lawyer?
- **Q4** ...using evidences **E1, E2, E3, E5**?

To compute the answer, we must consider that “the existence of a fact cannot be deduced by evidences, unless they are *severe, precise and coherent*” (article 192 of the Italian Code of Criminal Procedure). W.r.t. the fact “Valjean is the criminal with the red jacket”, **E1** alone is not even an evidence, as it does not link the criminal with Valjean; **E2** is severe but not precise, since there are explanations for Valjean’s presence of the fingerprint on the scooter’s mirror other than assuming he was riding it on the robbery day; **E3** is neither severe nor precise, since many persons besides Valjean can say “Jamunindi” and being born in Reggio Calabria does not necessarily imply to speak the local dialect; **E1 + E4** together represent a severe and precise evidence, since the time slots when the criminal with red jacket and Valjean were seen riding the same scooter were too close to allow a change in the scooter’s driver. Finally, **E5** is both severe and precise, as it proves that Valjean was in the crime scene at the time when the crime was

⁴https://archiviopdc.dirittopenaleuomo.org/upload/1445325933Trib_MI_Gennari.pdf, in Italian, last accessed on June 2020.

⁵The full description of the exercise is available, in Italian, here: <https://www.dropbox.com/s/8aof95p1q7ajx7k/Esercizio.pdf?dl=0>.

perpetrated.

The correct answers for the questions above are hence the following:

- **A1**, false: the availability of **E1**, **E2**, **E3** represent the real setting of the case of B.C., and indeed B.C. was acquitted by the judge Giuseppe Gennari on June 18th, 2015.
- **A2**, true: **E1 + E4** is severe and precise and it is coherently supported by **E2** and **E3**.
- **A3**, false: if Thenardier is unreliable, **E4** cannot be considered any longer and the situation becomes the same as in **A1**, where the only evidences that could be used were **E1**, **E2**, **E3**.
- **A4**, true: **E5** is severe and precise and it is coherently supported by the other evidences.

3 The Logical Judge and the Case of Valjean

The B.C. case offuscated by using different names and the code of the logical judge are implemented in SWISH⁶ and can be accessed at <https://swish.swi-prolog.org/p/casoValjean.pl>. Figure 1 shows a screenshot from that web page.

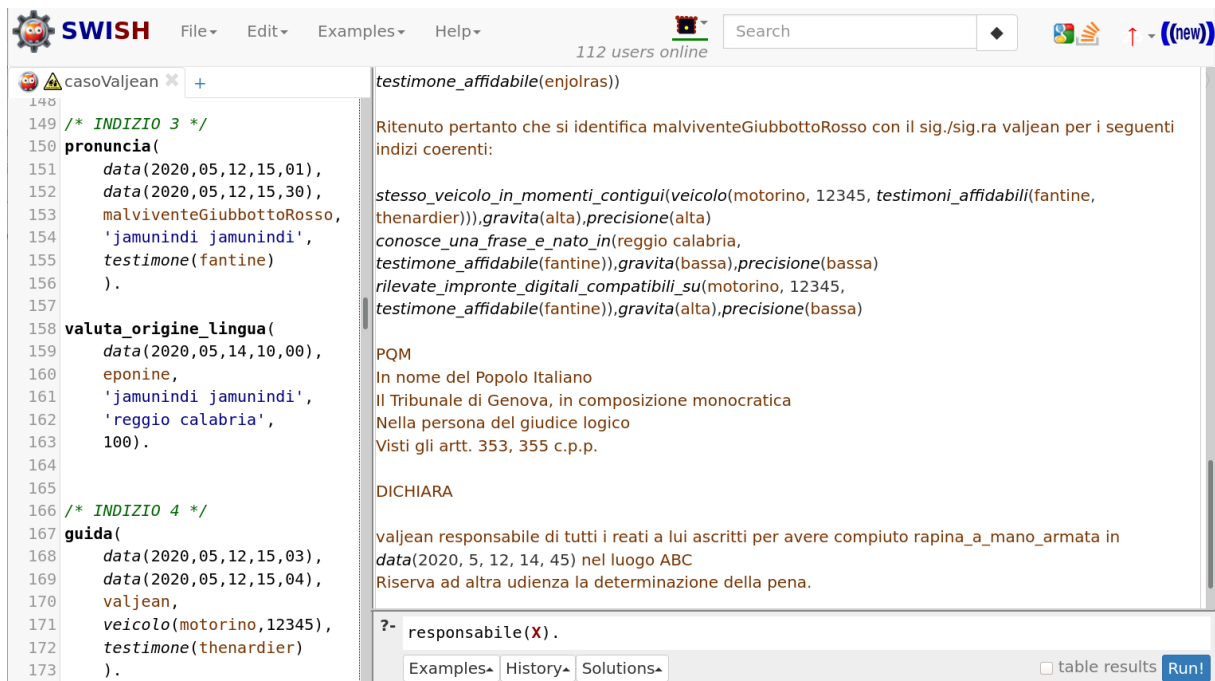


Figure 1: Screenshot of the logical judge on SWISH.

All the facts and evidences presented in Section 2 have been translated into Prolog facts, where predicates, functors, constants and variables are Italian words to allow the audience of the SSM course to immediately grasp their meaning. As an example, Figure 1, left pane, shows three facts that could be translated in English as

⁶SWISH is a web front-end for SWIProlog available from <https://swish.swi-prolog.org>, last accessed on June 2020.

```

/* EVIDENCE 3 */
utters(
  date(2020,05,12,15,01),
  date(2020,05,12,15,30),
  criminalInRedJacket,
  'jamunindi jamunindi',
  witness(fantine)).

words_origin_evaluation(
  date(2020,05,14,10,00),
  eponine,
  'jamunindi jamunindi',
  'reggio calabria',
  100).

/* EVIDENCE 4 */
drives(
  date(2020,05,12,15,03),
  date(2020,05,12,15,04),
  valjean,
  vehicle(scooter,12345),
  witness(thenardier)).

```

Information on Valjean, the witnesses, and the armed robbery are also modelled by Prolog facts like

```

born(
  date(1980,10,17,13,07),
  valjean,
  'reggio calabria').

commits(
  date(2020,05,12,14,45),
  criminalInRedJacket,
  armedRobbery,
  witness(enjolras)).

reliable(enjolras, hi).
reliable(fantine, hi).
reliable(thenardier, hi).

```

The reasoning mechanism is based on defining under which conditions one evidence backs up the fact that two individuals are the same one (for example, the fact that X and Y , with $X \setminus = Y$, were seen driving the same scooter in very close instants by two reliable witnesses is a highly severe and precise evidence that X and Y are same person), and collecting all those evidences, whose precision and severity may differ. If the evidences are at least two, supported by reliable witnesses, and at least one of them is severe and precise, we deduce that the two individuals are the same thanks to the following rule:

```

same_person(X, Y, Evidences) :-
  setof((Ev, severity(S), precision(P)),
    evidence_same_as(Ev, X, Y, severity(G), precision(P)),
    Evidences),
  length(Evidences, L), L > 1, member( (_, severity(hi), precision(hi)), Evidences).

```

The definition of responsible (X) is the following

```

responsibile(X) :-
  committed(Y, Date, Crime, Place, EvidCrimeCommitted),
  same_person(X, Y, EvidSamePerson),
  pretty_print(Date, X, Y, Crime, Place, EvidCrimeCommitted, EvidSamePerson).

```

where `committed` is defined by exploiting evidences and witnesses that support the fact that Y committed a crime and `pretty_print` prints the text shown in Figure 1, right pane. The printed text includes the motivations for the trial outcome and the formula that Italian judges utter to state their final decision.

By commenting and de-commenting evidences in the Prolog code and by changing the reliability of witnesses from `hi` to `lo`, we demonstrated to the audience that the different scenarios depicted in Section 2 can be simulated, with the logical judge answering **A1** to **A4** in the correct way. We also showed that changes to the Code of Criminal Procedure could be implemented by operating on the `same_person` predicate. For example, we might change $L > 1$ into $L > 0$ and state that only one evidence is enough, provided that it is precise and severe, which is actually the case of the Italian law. Or we might change it into $L > 3$ and remove `member((_, severity(hi), precision(hi)), Evidences)`, stating that at least four coherent evidences are considered as a proof, even if none is severe and precise.

The questions and remarks we received during and immediately after the working group activities showed that the topic raised the interest of the audience, and we felt we had successfully achieved our dissemination goals. To measure this feeling in a scientific way, a few weeks after the working group we

prepared a survey and invited the participants to answer a few questions. Figure 2 shows the results of this survey, aimed at assessing the background of the attendees and getting their feedback on the logical judge. Only 17 attendees answered the questionnaire, which was not mandatory and anonymous.

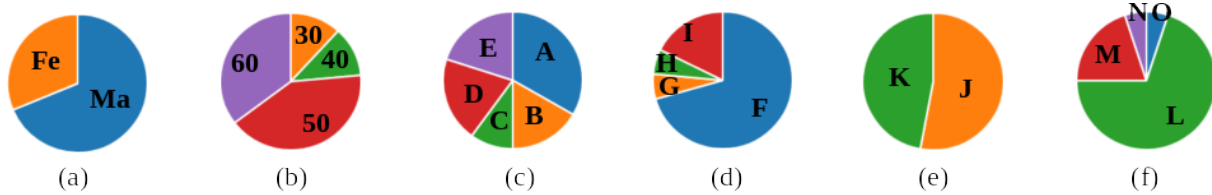


Figure 2: Figures from the post-course survey.

- Chart (a) shows the gender of those who participated in the survey: 11 males (69%) and 5 female (31%) (one participant did not answer).
- Chart (b) shows their age: 2 between 30 and 39 year old (12%), 2 between 40 and 49 (12%), 7 between 50 and 59 (41%), and 6 between 60 and 69 (35%).
- Chart (c) summarizes the participants' background on Artificial Intelligence, before following the SSM course; multiple answers to the questions about synonymy of popular locutions were allowed. 10 participants were aware that Artificial Intelligence and Algorithm are not synonym (slice A of the pie); 5 were aware that AI and Machine Learning are not (slice B); 3 knew that AI and Natural Language Processing are not (slice C); 6 knew that AI and Automated Reasoning are not (slice D). 6 participants stated that none of the above assertions was true for them (slice E). This last result is not surprising given the confusion about AI and the misuse of the term in non-technical (sometimes, also in technical) articles, but it is nevertheless worth some serious consideration that 6 participants out of 17 were not sure about the real meaning of popular words and terms like AI, algorithm, machine learning.
- Chart (d) presents the answers to our question on previous knowledge about logic programming languages. 12 participants did not know about their existence, but they appreciated their potential (71%, slice F), one was not aware of their existence and did not appreciate it, since he/she saw no applicability for them (6% , slice G); one was already aware of their existence (6%, slice H). 3 answered "Other".
- Chart (e) summarizes the answers about the possibility to have logical judges substituting, up to some extent, human judges. 9 participants felt that such approaches could offer a significant support to the judge in the future, but will never substitute her (53%, slice J), and 8 participants felt that the support given by such kind of approaches will be limited in the future (47%, slice K).
- Chart (f) presents the main feelings raised by logical judges. Multiple answers were allowed: 14 participants stated to be curious about their potential and limitations (slice L); 4 were looking for concrete results (slice M); one asserted to be worried about their potential raise (slice O). One answer was "Other".

Charts (e) and (f) support our feeling that we were able to raise the curiosity in the audience without neither scaring them, nor creating false expectations. We stress that our activities were carried out in less than two hours, and the background of the attendees was almost basic, as shown by Charts (c) and (d). One comment that emerged many times, both during the working group and as a free comments in the survey, is the need for judges to be exposed to a scientifically grounded and honest introduction to AI, in order to become informed and active players and decision-makers of their own future.

Acknowledgements. This paper is based upon work from COST Action DigForASP, supported by COST (European Cooperation in Science and Technology).

References

- [1] Christoph Benz Müller, Xavier Parent & Leendert van der Torre (2020): *Designing normative theories for ethical and legal reasoning: LogiKEy framework, methodology, and tool support*. *Artificial Intelligence*, p. 103348, doi:<https://doi.org/10.1016/j.artint.2020.103348>.
- [2] Richard Berk (2019): *Machine learning risk assessments in criminal justice settings*. Springer, doi:<https://doi.org/10.1007/978-3-030-02272-3>.
- [3] Marc A. Borrelli (1990): *Prolog and the Law: Using Expert Systems to Perform Legal Analysis in the United Kingdom*. *Softw. Law J.* 3(4), p. 687–715, doi:<https://dl.acm.org/doi/10.5555/91456.91463>.
- [4] Nuttawit Butsara, Panchan Athonthitichot & Pichai Jodpimai (2019): *Predicting Recidivism to Drug Distribution using Machine Learning Techniques*. In: *2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE)*, IEEE, pp. 1–5, doi:<https://doi.org/10.1109/ICTKE47035.2019.8966834>.
- [5] Roberta Calegari, Giuseppe Contissa, Francesca Lagioia, Andrea Omicini & Giovanni Sartor (2019): *De-feasible Systems in Legal Reasoning: A Comparative Assessment*. In: *Legal Knowledge and Information Systems - JURIX 2019, Frontiers in Artificial Intelligence and Applications 322*, IOS Press, pp. 169–174, doi:<https://doi.org/10.3233/FAIA190320>.
- [6] Stefania Costantini, Giovanni De Gasperis & Raffaele Olivieri (2019): *Digital forensics and investigations meet artificial intelligence*. *Annals of Mathematics and Artificial Intelligence* 86(1-3), pp. 193–229, doi:<https://doi.org/10.1007/s10472-019-09632-y>.
- [7] Mirna El Ghosh, Hala Naja, Habib Abdulrab & Mohamad Khalil (2017): *Towards a legal rule-based system grounded on the integration of criminal domain ontology and rules*. *Procedia CS* 112, pp. 632–642, doi:<https://doi.org/10.1016/j.procs.2017.08.109>.
- [8] Guido Governatori, Antonino Rotolo & Régis Riveret (2018): *A Deontic Argumentation Framework Based on Deontic Defeasible Logic*. In: *International Conference on Principles and Practice of Multi-Agent Systems*, Springer, pp. 484–492, doi:https://doi.org/10.1007/978-3-030-03098-8_33.
- [9] Allen Hustler (1982): *Programming law in logic*. Research report CS-82-13, Department of Computer Science, University of Waterloo.
- [10] Han-Wei Liu, Ching-Fu Lin & Yu-Jie Chen (2019): *Beyond State v Loomis: artificial intelligence, government algorithmization and accountability*. *International Journal of Law and Information Technology* 27(2), pp. 122–141, doi:<https://doi.org/10.1093/ijlit/ez001>.
- [11] Masha Medvedeva, Michel Vols & Martijn Wieling (2020): *Using machine learning to predict decisions of the European Court of Human Rights*. *Artif. Intell. Law* 28(2), pp. 237–266, doi:<https://doi.org/10.1007/s10506-019-09255-y>.
- [12] Jeffrey Alan Meldman (1975): *A preliminary study in computer-aided legal analysis*. Ph.D. thesis, Massachusetts Institute of Technology.
- [13] William Lloyd Prosser (1941): *Handbook of the Law of Torts*. *Louisiana Law Review* 4.
- [14] Cynthia Rudin (2019): *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*. *Nature Machine Intelligence* 1(5), pp. 206–215, doi:<https://doi.org/10.1038/s42256-019-0048-x>.
- [15] Marek J. Sergot, Fariba Sadri, Robert A. Kowalski, F. Kriwaczek, Peter Hammond & H. T. Cory (1986): *The British Nationality Act as a Logic Program*. *Commun. ACM* 29(5), pp. 370–386, doi:<https://doi.org/10.1145/5689.5920>.
- [16] Songül Tolan, Marius Miron, Emilia Gómez & Carlos Castillo (2019): *Why machine learning may lead to unfairness: Evidence from risk assessment for juvenile justice in Catalonia*. In: *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, pp. 83–92, doi:<https://doi.org/10.1145/3322640.3326705>.
- [17] Alessandro Traversi (2014): *La difesa penale. Tecniche argomentative e oratorie*. Giuffrè Editore.

- [18] Douglas Walton (2018): *Legal Reasoning and Argumentation*. In: *Handbook of Legal Reasoning and Argumentation*, Springer, pp. 47–75, doi:https://doi.org/10.1007/978-90-481-9452-0_3.
- [19] Hajime Yoshino, S Kagayama, S Ohta, M Kitahara, H Kondoh, M Nakakawaji, K Ishimaru & S Takao (1986): *Legal expert system–LES-2*. In: *5th Conference on Logic Programming*, Springer, pp. 34–45, doi:<https://dl.acm.org/doi/abs/10.5555/647633.732885>.