Representing Norms with Extended Argumentation Frameworks

Trevor Bench-Capon
Department of Computer Science, University of Liverpool
Sanjay Modgil
Department of Computer Science, King's College, London

April 11, 2019

Abstract

This paper discusses reasoning about norms using Extended Argumentation Frameworks. We start by exploring Rawl's *Theory of Justice* through the use of computational argumentation. Scenarios are modelled using Action Based Alternating Transition Systems with Values, and the resulting arguments evaluated in an Extended Argumentation Framework, which permits arguments to attack attacks as well as arguments, and so provides a natural way of representing exceptions. We also offer an interpretation of the promotion of values as the fulfillment of needs, and use Maslow's hierarchy to choose between these needs when reasoning about Rawl's initial position. Following our consideration of how norms are *justified*, we consider their application. Our representation of norms in EAFs enables us to offer a new account of permissions and exceptions. We illustrate our approach with an example relating to some US Supreme Court Fourth Amendment cases.

1 Introduction

In [11], [7] and [12] a discussion of the relation between norms and value based reasoning[1] was initiated. State transition diagrams are commonly used in multi-agent systems to represent the actions available to the agents, and the norms applicable to them (e.g. [6]). One common way of representing prohibitions (e.g. [1]) is to remove the transition representing the forbidden action. But this makes norms implicit and, worse, prevents any possibility of violation, which is central to any discussion of norms [22]. Essentially this way or representing norms reduces *should not* to *can not*. Our approach is to use the transition diagrams to generate value based arguments, as in [2]. In this way norms can be seen to emerge from the process of practical reasoning, reasoning about what actions should be chosen, and can then be represented as arguments for and

---

[1]For those unfamiliar with value based reasoning, the underpinning semantic structures (Action-based Alternating Transition Systems with Values (AATS+V) [2]), Value Based Argumentation Frameworks (VAF) [4] and Extended Argumentation Frameworks (EAF) [26] are defined in the Appendix.

against the actions they concern. We use Extended Argumentation Frameworks (EAF) [26] to represent and evaluate the relevant arguments, to enable reasoning about both what the norms should be, and, once the norms have emerged, how they should be applied. EAFs not only enable a more expressive and elegant representation of preferences, but also provide a very natural representation of exceptions, which are of great importance in any discussion of legal norms.

The paper is structured as follows. Section 2 introduces a basic scenario which will be the basis of our discussion of the emergence of norms. Section 3 discusses how norms can emerge from this scenario by applying value based reasoning to Rawl's notion of justice as fairness [30]. We then relate the values used to Maslov's hierarchy of needs [25] so as to give a more objective basis to the values used and the preferences between them. This section also considers how best to represent the argumentation, proposing EAFs as an improvement on standard VAFs. Thus far norms have been a *product* of need based practical reasoning about what agents should do, and so norms are not explicitly represented. In our view, a norm can be seen as effectively encapsulating some reasoning from first principles, so that in future that reasoning can be applied by the application of the norm, without need to reconsider its justification. Section 4 discusses the representation of norms as arguments, and the need to accommodate exceptions, in situations in which it is accepted that the norm can be violated. Section 5 illustrates the advantages of our approach by applying it to a set of US 4th Amendment cases, much discussed in AI and Law. Finally Section 6 offers some concluding remarks. An appendix contains the relevant formal definitions.

2 Basic Scenario

To describe our approach we will use a scenario which was introduced in [14] and was used to discuss norms in [11], [7] and [12]. The scenario could serve as the background either for the fable of *The Ant and the Grasshopper*[2] or the parable of *The Prodigal Son*[3], depending on the choices made by the agents in the scenario. In the stories one person (the grasshopper or the son) idles away the summer, and so has no food when winter comes. Working produces a surplus of food, so that the ant (father) could give food to the grasshopper (son) without themselves starving. In the fable the ant refuses the grasshopper, arguing that she has only herself to blame, but in the parable the father welcomes the son and gives him food. Figure 1a shows the transition diagram used in [14]. Each state is represented by a 5-tuple which indicates: whether the ant/father is alive, whether the grasshopper/son is alive, how much food the ant/father has (2 is surplus, 1 is enough, 0 is none), how much food the grasshopper/son has and whether it summer (1) or winter(0). In summer their options are to work (w) or play (p), giving rise to four joint actions. If a person has no food, he can only ask for food (a). If a person has food in winter he may give food to the other (g) refuse to give (r) or, if no request is made, just eat (e). Finally if a person has eaten and still has food they may feast (f). Uneaten food is assumed to spoil by the end of winter. The transitions are labelled

---

[2]One of Aesop's Fables, numbered 373 in the Perry Index.
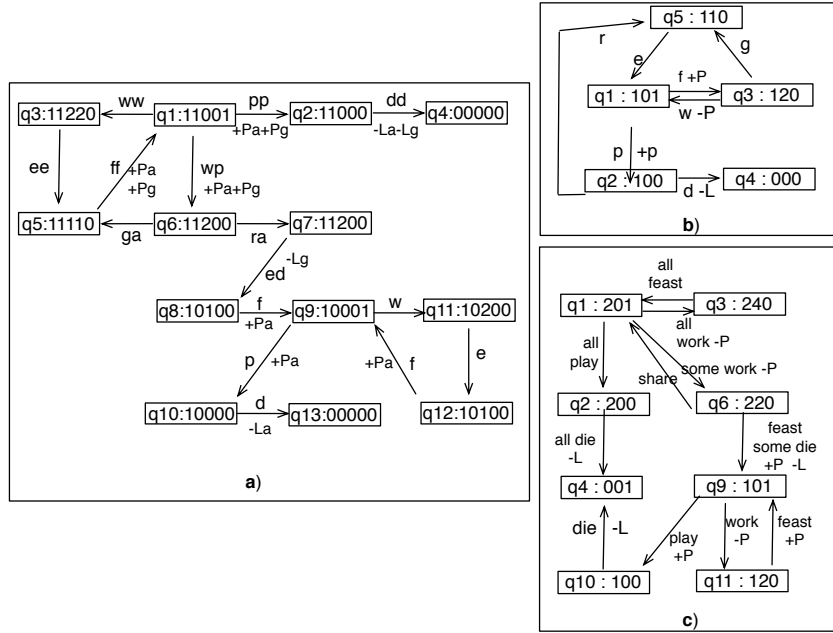[3]Luke 15:11-32

Figure 1: Key: w = work, p = play, a = ask, g = give, r =refuse, e = eat, f = feast d =die.
a) AATS giving Bird's Eye View. The same AATS+V is used for both the fable and
the parable. Joint actions are ant/father, grasshopper/son. States are: ant/father alive,
grasshopper/son alive, ant/father has food. grasshopper/son has food, summer/winter.
b) AATS for the Individual Perspective. **c) AATS for the Community Perspective**.

with the joint actions composed from the action of the ant/father and the grasshopper/son: e.g. *wp* is ant works and grasshopper plays. Finally the transitions are also labelled with the values promoted (+) and demoted (-). The values used are pleasure and life. These are relevatised to the person enjoying the pleasure or the life: Pa/La is pleasure/life for the ant/father and Pg/Lg is pleasure/life for the grasshopper/son.

Figure 1a shows the "birds" eye view, as given in [14] and used in [7] and [11]. This enables us to see who does what, and in respect of whom values are promoted and demoted. In this paper we will also consider the scenario from the perspectives of an individual (Figure 1b) and the community (Figure 1c). The individual perspective considers only the actions of an individual, and does not distinguish between the ant and the grasshopper. This means that different outcomes of an action cannot be explained by the actions of other agents. Whereas the individual perspective sees only the actions of one agent, the community perspective sees only the joint actions without distinguishing individual contributions, and does not distinguish individuals in the states. Thus, for example, the community perspective considers only the number of living members, whereas the bird's eye view records which individuals are alive.

There are two choices to be made. One, made in q_1, is between work and play.

From the individual perspective, working is a safe choice, ensuring that the individual will be alive to work next summer. A worker will reach q_3 where he can choose to share his surplus food, or indulge in feasting. An agent who chooses play in q_1, reaches q_2 where there is no choice: food may be received, or not, and this external choice will determine whether the agent lives or dies. From the community perspective, if all work the population is maintained: otherwise there is a risk of population decrease.

In [11] and [7] it was argued that, from the Bird's Eye view, both choices in the two choice situations are legitimate, and individuals of different temperaments may choose either of the options. There does not seem to be a need to label the choices as moral or immoral: there is no moral reason to prefer a long and tedious life over a short and happy one, nor any moral obligation to prefer the life of another to one's own. Neither is it necessarily morally correct to allow oneself to be exploited by free loaders, rather than punish them: indeed the reverse may be true [24]. None the less, as we shall see, norms are still emerge from this scenario, using the principles found in [31].

3 Value Based Reasoning and Rawls' Theory of Justice

If we take the bird's eye view as representing the initial situation of various social contract thinkers [4], the effect of the perspectives is to introduce what Rawls [31] terms a *veil of ignorance*. Whereas in the bird's eye view an agent can identify itself distinct from others, the other two perspectives do not allow this. In the individual perspective anonymity comes from not knowing the role that the agent will play, and in the community perspective anonymity comes from aggregation. Rawls' intuition is simple: that if there will be two different outcomes to be decided by an agent, it will attempt to be as fair as possible if it does not know which of the outcomes will apply to itself. A common example is when two children share a cake, one cutting and the other choosing: the cutter will strive to divide the cake as equally as possible since she may receive either of the two pieces. Rawls uses what he calls the *difference principle* [34] which means that the chosen strategy will be the option in which the worst off does best (since no one knows whether they will be part of this worst off group). We can apply this to the well known games discussed in [9] and [3].

In Prisoner's Dilemma [33] there are three scenarios: both cooperate, both defect, and one cooperates and one defects. The outcomes for the worst off are 3, 1 and 0 respectively. Thus mutual cooperation will be chosen on the difference principle (rather than the Nash Equilibrium [5] of mutual defection) since the agent should not risk being in a cooperate-defect situation, since although the defector gains 5, it may itself be the co-operator and receive 0. Similarly, in the Ultimatum [19] and Dictator [18] games, the chosen strategy will be equal division, rather than the game theoretic recommendations of giving zero in the Dictator game and the minimum amount in the

---

[4]Called the *state of nature* by Hobbes [21], the *initial bargaining position* by Gauthier [17] and the *original position* by Rawls [31].

[5]If no player can benefit by changing strategies while the other players keep theirs unchanged, then the current set of strategy choices and their corresponding payoffs constitutes a Nash equilibrium [29]. As such, the Nash Equilibrium strategy is often taken to be the most plausible strategy.

Ultimatum game, since the agent does not know which of the roles it will play, and so should prefer to ensure that it gets half rather than risk getting nothing.

In the fable we have four scenarios: 1) both work; 2) both play; 3) one works and gives, and; 4) one works and refuses. In 2) and 4) the worse off dies, and in both 1) and 3) the worse off works, but in 3) the worse off gives some of the fruits of his labour to the other. The best outcome for the worse off is, therefore, scenario 1) in which both work, and so according to the difference principle this should be the rational choice. Otherwise the agent may find that it is the one which works to support the free loader, and play represents an unacceptable risk. This is clearly the best choice from the community perspective since it is desirable that all work if the safety of the members of the community is to be ensured, and the population maintained. In [7] it was considered how imbalances of power and wealth could lead to different social organisations, since the powerful and the wealthy could play in the summer and then obtain their food by force or purchase. Rawls is attempting to rule out these organisations[6] by using his *veil of ignorance*: he argues that if the agents do not know who will have the power and money they will prefer an equal society. Moreover, Rawls is arguing for a democracy, where the norms are not established by the powerful and the wealthy but by a consensus. As expressed in [34], the aim of [31] is to envision a society of free citizens holding equal basic rights cooperating within an egalitarian economic system. Thus, following Rawls' notion of *justice as fairness*, we would expect the following norm to emerge:

- N1: It is obligatory to work in the summer

However, even if all the citizens recognise this as the desirable norm, this will not prevent particular agents attempting to get a personal benefit by violating the norm. In general people are happy to accept that people should pay their taxes, because all benefit from public services. None the less many people will seek an advantage by not paying their own taxes. For this reason, it is necessary that violations of the norm be punished, so that there is no advantage in violation. Thus we also require

- N2: It is obligatory to punish a non-worker by refusing food

Without such a sanction there is no deterrent to violation and normative collapse is inevitable, as demonstrated in simulations such as [24].

3.1 Variant Transition Diagrams

.

Figure 1a represents a very basic model of the situation and assumes an equal starting point. In [7] and [11] it was shown how the arguments that could be generated from the AATS model depended crucially on what was included in the model, and how it was labelled. In turn, the norms that could be articulated depended crucially on the arguments that could be generated.

---

[6]His first principle of justice as fairness is " Each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all".

Thus, by including more in the state, it is possible to represent the imbalances of power and money alluded to above, allowing norms to govern the proper use of power and wealth, and to regulate monetary transactions, including where payment is in arrears. By representing the history, so that one can know how many times the offence has been committed one can also allow a certain degree of tolerance. This might take the form of not punishing a first offence, so that the first time round the worker could give food, as in the parable, but deny it to punish any subsequent offence. Although simulations such as [24] demonstrate that it is necessary for transgressions to be punished if normative collapse is to be avoided, others such as [23] have suggested that a degree of tolerance before deciding to punish may be beneficial, enabling mutual cooperation (or a system of mutually advantageous turn taking) to be established, with the accompanying long term benefits.

Similarly the way actions are labelled can have an important impact. If there is an imbalance of power the grasshopper can demand, rather than request food, and the ant has no option to refuse. If there is an imbalance of wealth the grasshopper can offer to buy rather than beg, and the ant can sell or not as it chooses. Moreover, whereas here any non-working activity is called "play", in [7] a number of distinctions were made between worthwhile and idle play, and between activities that the rest of the community might be prepared to pay for (such as entertainment or law enforcement) and those they were not prepared to pay for. These variations will be discussed further in section 4.

Finally, the use of different values to label the transitions will change the flavour of the argument. Presented as a choice between the life of the grasshopper and the pleasure of the ant, refusal seems hard hearted if not immoral, whereas presented in terms of punishment so that refusal promotes Justice and agreement promotes Mercy, the choice to refuse can even seem virtuous (especially given that too much Mercy will eventually lead to normative collapse).

In this paper we are especially interested in how norms are justified using value based argumentation. Such justification will require a principled basis for the choice of values and preferences between them. Hence we begin by examining more closely how the values that are used to label the transitions might be chosen.

3.2 Labelling the Transition Diagram with Values

The original labelling from [14] used the two values of Life and Pleasure (both of which can be relativised to the particular agent when using the bird's-eye view). These are quite fundamental and non-specific, but other work using value-based reasoning has used many quite disparate things as values. Thus in [11] we find the highly domain specific values of Progress and Safety used to reason about Traffic Law, whereas in the more general travel situation of [8] we find values such as Comfort and Speed. In discussions of US Trade Secret Law such as [15], the values were taken to be various desiderata in striking a balance between competition and innovation in a modern industrial society. A consideration of the literature would seem to suggest that values are determined pragmatically in a rather ad hoc fashion, according to the specific problem. We need to introduce a bit more principle into value selection, especially if we want to

apply it to Rawls' *original position*.

To achieve this we think about actions being chosen based not on the *values promoted* but on the *needs fulfilled*. The intuition is that for the basic situations which can serve as Rawl's original position we need to consider aspects more fundamental and universal than the matters of taste and fashion, or the specific subjective concerns, that values are often used to represent. One source of such fundamental values is Maslow's Hierarchy of Needs [25]. In that theory, human needs are organised into a five-level hierarchy.

1. *Biological and Physiological needs* - air, food, drink, shelter, warmth, sex, sleep.

2. *Safety needs* - protection from elements, security, order, law, stability, freedom from fear.

3. *Love and belongingness needs* - friendship, intimacy, trust and acceptance, receiving and giving affection and love. Affiliating, being part of a group (family, friends, work).

4. *Esteem needs* - achievement, mastery, independence, status, dominance, prestige, self-respect, respect from others.

5. *Self-Actualization* needs - realizing personal potential, self-fulfillment, seeking personal growth and peak experiences.

From this hierarchy we can locate Life at level 1 (specifically we are thinking about the need for food, but this is really just representative of these other basic needs), and Pleasure deriving from play at Level 5. The pleasure derived from consuming the surplus in community feasts (Harvest, Christmas, Carnival) will address level 3 needs. Seen in this way the value preferences expressed by N1 and N2 can be seen as preferring more fundamental needs to higher level needs.

3.3   Representing the Argumentation

The traditional Value-Based Argumentation Framework (VAF) [4] (see Appendix for formal details) is shown in Figure 2a. Here we have two arguments. Arg1 is an argument claiming one should work, based on promoting the value of life ($+L$), whereas Arg2 claims one should play, based on promoting the value of play ($+P$). The conflict is resolved by an 'audience's' preference between the two values. Assuming an audience that prioritises the value of play over life, *Arg*1 now no longer defeats *Arg*2; only *Arg*2 defeats *Arg*1 and *Arg*2 is justified. On the other hand, assuming an audience prioritising life over play, *Arg*2 no longer defeats *Arg*1; only *Arg*1 defeats *Arg*2 and *Arg*1 is justified.

One might consider encoding the rationales for (i.e., arguments justifying) the contrary audience preferences in the framework, as shown in Figure 2b. Then choosing the argument $P > L$ ($L > P$) claiming that play (life) is a more important value than life (play), in the absence of an argument for the contrary preference $L > P$ ($P > L$), results in *Arg*1 (*Arg*2) being defeated, and *Arg*2 (*Arg*1) being justified. Although this works pragmatically, it is not strictly correct: the preference does not really attack the
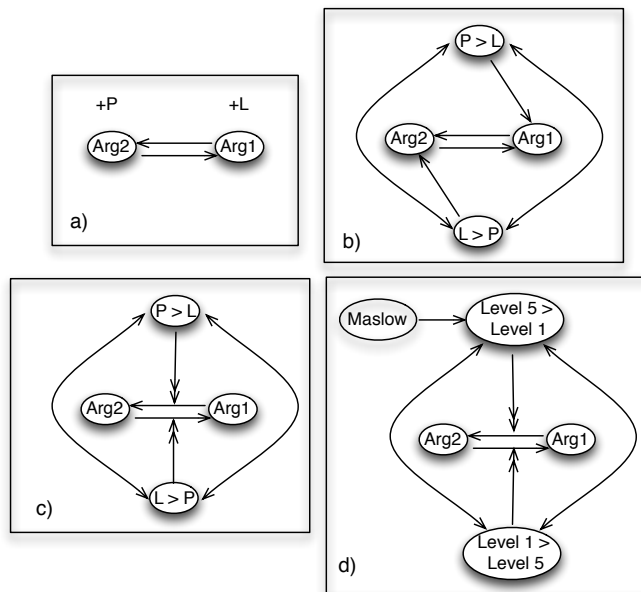
Figure 2: a) Basic VAF for q_1. b) Basic VAF for q_1 with arguments expressing value preferences. c) Basic EAF for q_1. d) EAF for q_1 with Maslow.

weaker argument, which remains a perfectly good argument, even though it cannot be *accepted* in the particular situation by an audience with those particular preferences. A much better representation is given by the use of Extended Argumentation Frameworks (EAFs) [26], that extend the classical Dung argumentation frameworks [16] to incorporate arguments encoding rationales for (possibly conflicting) preferences and/or value orderings. In EAFs, such arguments then attack attacks, as shown in Figure 2c. Now, as it made clear in [27] and[28], the preference attacks not the *validity* of the weaker argument, but the *statement that it defeats* the preferred argument[7]. EAFs are formally defined in the Appendix. Intuitively, if an argument X attacks the attack from Y to Z, then X invalidates the use of Y as a counter argument to Z, e.g., because X expresses that Z is stronger than (preferred to) Y, or X expresses exceptional conditions under which Y does not counter argue Z.

However Figure 2c still permits two possibilities. One set of justified (acceptable) arguments includes $Arg2$ and $P > L$, while the other includes $Arg1$ and $L > P$. As yet there are no arguments arbitrating in favour of one preference over the other. But if we now make use of Maslow, and replace the values with the corresponding level of need, and take Maslow's hierarchy as an argument suggesting that lower level needs need to be satisfied before we can explore higher level needs, we get the EAF in Figure 2d, in which the only set of justified arguments includes $Arg1$ advocating work, as well as $Level1 > Level5$ and $Maslow$. We therefore have a justification for N1.

Now consider the situation in q_6 in which one agent (the ant) has complied with the norm and one (the grasshopper) has not. Now Arg1 is to give the grasshopper food and Arg2 is to refuse food. Note that now the arguments are relevant only to the ant and the pleasure is the ant's and the life is the grasshopper's. This complicates matters. Maslow would tell us that if both values relate to the ant, life should be preferred over pleasure, but here we have a level 3 need of the ant conflicting with a level 1 need of the grasshopper. For this reason we perhaps need to think beyond these needs and think more in terms of refusal being punishment for violation of N1, and so promoting Justice. But even so, we are contrasting this with Compassion or Mercy, both of which might be held to be preferable to Justice[8].

Let us however consider the situation of q_6 from the position of the ant, considering only the needs of the ant. In q_6 all the level 1 needs of the ant are satisfied. Where everyone complies (q_5), the surplus food is used in community feasting and celebration, so fulfilling certain level 3 (belongingness) needs (the social need for unity). However belongingness is recognised to have a number of forms: we belong to a family, a circle of friends, a group of work colleagues, and a local community. Thus in q_6 the ant might prefer to meet the needs of the grasshopper if they have one of the closer relationships: family, friendship, or collegiate relationship, rather than use the surplus to participate in the wider community activity. This pull will differ between societies, and between

---

[7]Suppose in Figure 2b one chooses $P > L$ so that $Arg1$ is defeated. Suppose then that an additional argument undermines (attacks) $Arg2$ by claiming that the required conditions enabling play are not present. It then remains the case that $Arg1$ is defeated, whereas one would now want this argument for working to be justified. In the EAF representation in Figure 2c, attacking $Arg2$ *does now* result in $Arg1$ being justified (see Appendix).

[8]A justification for the preference for Justice could be seen in terms of society itself having needs such as sustainability (Level 1), order (Level 2), unity (Level 3) etc.

different members of a given society. But given the existence of the norms, there are also level 2 needs to be considered: the needs for security, law, justice and order require that violations be punished. Thus the preference for Justice over Compassion can be seen in terms of the Maslow hierarchy, with a preference for level 2 needs over level 3 needs. This would provide a needs based justification for N2, so that we have explained both our norms in terms of preferences for lower level needs over higher level needs.

4 Representing Norms and Exceptions

Once we have decided what the norms should be, we need to represent them so that they can directly influence the choices of agents, rather than requiring them to derive the norms from first principles every time. The arguments used in the frameworks so far have been of the form *citizens should perform a particular action in order to fulfill some need*. Promulgation of norms are essentially the actions available to the state. The actions available to the citizens are (in State Transition terms) to follow transitions. In many multi-agent systems (MAS) which use norms (such as [1]) norms are represented by removing the transitions corresponding to prohibitions. Obligations (if represented) are represented by removing all transitions from a state *except* the one that is obligated. Permissions seem rather mysterious on this approach, since the existence of the transition makes them permissible without a norm[9]. The real problem, however, with removing transitions in this way is that it removes any possibility of violation, while the point of representing norms is precisely so that violations can be reasoned about [22].

In this spirit we first represent norms as attacks on prohibited actions, as shown in Figure 3a. Note that we have replaced the reasons for introducing the norm, whether based on Rawl's difference principle, or using value preferences based on the Maslow hierarchy, by the argument *Norm*, expressing that *play is forbidden*. Once the norm has been introduced it *replaces* the arguments which gave rise to it: citizens are expected to accept the norm, rather than dispute its justification. The situation is similar to that of Hare's account of moral principles in [20]. Hare argues that, in a novel situation, we need to reason from first principles, but having arrived at a conclusion we "crystallise it into a not too specific or detailed form, so that its salient features may stand out and serve us again in a like situation without so much thought." In other words, having developed the moral principle or norm it can be applied without reference to the arguments which originally gave rise to it.

The problem with the representation of Figure 3a, as noted in [10] and [11], is that norms are made in the knowledge that they will be violated, and in certain circumstances, *should* be violated. EAFs provide a very natural way to represent this, by having the exception attack the attack originating from the norm, as shown in Figure 3b. Thus the exception means that either action is permissable, and agents can again choose according to their value preferences. Note that an exception removes the com-

---

[9]Although deontic logic has always focussed on obligation and permission, in practice prohibitions are the more usual way of expressing norms. Nine of the Ten Commandments found in Exodus 20:1-17 are expressed as prohibitions. This focus on forbidding actions (even when the aim is to avoid undesirable states as in [35]) is adopted also in MAS.
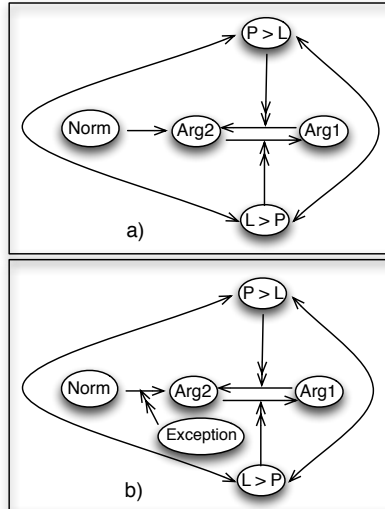
Figure 3: a) EAF with norm. b) EAF with norm and exception.

pulsion on one action, but does not compel the other. Compulsion would require the exception to also attack the argument for the action enjoined by the norm. Now let us consider the form that the exceptions might take.

Considering Norm 2, which is motivated by the need to punish violations of Norm 1, the exceptions will be the cases where the agents which *prima facie* should be punished are considered to be exempt from N1. Thus the exception arguments will be based on various reasons why it may be thought that N1 does not apply to the agent concerned. These exceptions will need to be coupled with a norm obligating the giving of food to the people concerned, so that their freedom not to work[10] is not accompanied by an unacceptable risk, and the survival of those exempt from work is assured. Candidates for such exceptions abound (although they would require a richer AATS+V than the one shown in Figure 1a to be expressible):

- Those who cannot work. What counts as incapacity varies from society to society, but most societies recognise that some people will be unable to work, at some stages of their lives. Typical reasons are age (the young and the old are often exempt), illness, and the need to provide child care.

- Many societies support some kind of historian, storyteller, praise singer, poet and/or musician, such as the bards of ancient Greece, the minstrels of medieval

---

[10]In this section "work" will be used exclusively to mean labour to produce food, even though some of the exceptions will require that exemption from this kind of work will involve some alternative useful (paid) employment, such as law enforcement, entertainment and building, to be undertaken instead. Although these activities may be considered work, they would typically be considered preferable to labouring to produce food, and are often chosen voluntarily, and often offer the satisfaction of higher level needs.

Europe, or the griots of West Africa. These people provide an important source of entertainment and pleasure to the other members of the community, but they also fulfill a number of level three needs, bringing the community together, articulating its history and ideals. The role is desirable to the bard since it allows the bard to fulfill a number of level 4 esteem needs, providing opportunities for mastery, prestige and respect from others. This pattern in which the exception fulfills level 3 needs for the community and level 4 needs for the excepted individual is typical of several exceptions.

- People responsible for Law Enforcement. As noted in [24], since punishment has a cost, and the need to punish may fall unevenly, it is quite natural for a society to pay an agent to enforce the norms and punish violators on their behalf. By spreading the risk of needing to punish and so acting as a form of insurance (increasing the likelihood that transgressors will be punished), independent law enforcers satisfy some of the level two needs such as security and the need for law. Transfer of punishment from vigilante groups to the state is a sign of growing civilisation and is an important part of the move from social convention to a proper legal system.

- As well as law enforcement, there are other public services which can benefit from dedicated public officials removed from the workforce for that purpose. For example people responsible for the distribution of alms, and various community leadership functions may emerge.

- In many societies, there is a religion which is supported from surplus production enabling the existence of dedicated priests and monks. This is perhaps more controversial than some of the other groups, since different states have enforced support for religion, tolerated the support of religion, or actively opposed it. Maslow does not himself mention religion in his discussion of belonging needs.

- Other forms of civic pride, such as builders. Some societies express themselves through community building projects, such as the pyramids, cathedrals and the Millennium Dome in Greenwich. This again reinforces level 3 needs such as belonging for the citizens, and level 4 needs for the architects and craftsmen.

- In some societies a leisured class has developed, exempt from working. Such a class was discussed at length in [7], but is something which the democratically minded Rawls would hope to avoid through his "veil of ignorance".

These exceptions enable the fulfillment of higher level needs. With the possible exception of law enforcement officers and a leisured class, these candidate exceptions belong (from the citizen's perspective) to level 3: they represent different ways of enabling a sense of community. There is therefore considerable scope for choice and variation in selecting exceptions, and determining their extent and competing merits. Support for these activities will require a food surplus. Assuming that the surplus is insufficient to allow all potential exceptions, so that the possible exceptions are competing with each other, it is here that the *values* of particular societies reveal themselves. Thus a compassionate society will allow a broader definition of incapable of working

than an austere one. A more invidualistic society might not wish to support a bard, whereas bard-led events will be an important part of a more collectivist society (see for example the decline of working men's clubs in the UK as strongly collective groups such as miners and other industrial workers have declined in the UK). Most societies will have some kind of law enforcement and leadership, but will vary in how authoritarian, libertarian and self-reliant they are, It has already been noted that societies will vary in their piety and devoutness. Finally, societies may be more or less egalitarian, and more or less deferential towards inherited status. The prevailing values of a community influence how the surplus will be used and hence what will count as an exception to N1 in that particular society.

The above values are related to the tastes and predilections of the citizens. As noted above in footnote 7, however, communities themselves have needs of their own, which can be seen as forming a Maslow-like hierarchy. Thus at the most basic level the state wishes to maintain its population, and perhaps grow it. At level 2 the state will wish its people to be secure and free from anxiety. But as we rise the hierarchy different expressions become possible: a state may wish to increase its wealth, and may choose to do this through trade or military force. Thus either a merchant class or a military class may develop. Similarly, power may be pursued in either its hard form (as exemplified by ancient Sparta), or its soft form (as exemplified by ancient Athens). Thus in some societies writing poetry may be a source of prestige and respect, and in others a source of derision and scorn. Similarly being part of the military may or may not enjoy the respect of others depending on their values.

In Figure 3b, the effect of the exception is to accommodate choices determined by the value preferences of the society involved. But in other cases the exception will need to rule out other arguments as well as the applicability of the norms. Thus in Figure 2d, the societal needs served by the exceptions will attack the attack originating from the Maslow argument, indicating that we are not using Maslow's hierarchy to make this choice. If an exception were to be incorporated in the law as a fully fledged norm in its own right, free-standing rather than an exception to another norm, it will attack the argument to work in respect of those exempt classes of people, rather than attacking only the attack from the original norm.

4.1 Permissions

Permissions have always seemed mysterious in Multi-Agent Systems: prohibitions are represented by removing a transition, and so permissions are represented by retaining it: but then what does permitting an action add? Our framework suggests an answer. We have already seen that three different attitudes on the part of the state towards an exception are possible; the sharpest example being the treatment of religion. The exception may be obligatory, voluntary, or forbidden. Thus in the UK, the right to receive tithes was granted by King Ethelwulf in 855, greatly modified by Henry VII as part of of the dissolution of the monasteries in the sixteenth century, which abrogated many of the rights to the crown, and made entirely voluntary by the Tithe Commutation Act of 1836. The USSR and China in the twentieth century provide examples where payments to religious organisations have been forbidden. Thus all three deontic modalities have

been applied to this exception. How will they manifest themselves in our representation?

- An obligatory exception is presented as in Figure 3b: it is an argument without attackers. In an even stronger form it will also attack the argument for the action enjoined by the norm (Arg1 in Figure 3b).

- A permitted exception attacks, and is attacked by arguments for competing exceptions. Thus there may be alternative sets of justified (acceptable) arguments that include the competing exceptions, and it is then the value preferences of individual citizens that determines which exceptions take priority. Permissions to disregard N1 also appear as exceptions to N2.

- A forbidden exception is attacked by, and does not itself attack, an argument that the exception is not permitted (which may be a norm). Thus the argument expressing the exception does not appear in any set of acceptable arguments, and its attack on the attack from the norm will be ineffective.

.

Exceptions may also attack different arguments depending on their normative status. A permissive exception will allow the choice to be based on a preference, whereas an obligatory exception can also supply an attack on the argument for the action which is not required given the exception. If the exception is forbidden, it will be attacked by an argument based on the prohibiting norm.

4.2   To Whom Should the Exceptions Apply?

Some exceptions relate to meeting the lowest level needs. If a person is incapable of meeting their needs themselves, they must look to others to meet them. But what counts as incapacity varies. The age at which children are considered capable of paid employment has steadily risen over the ages, and legislation exists in may countries to control the age at which children may work, and the number of hours they may work for. Similarly, when people are entitled to retire has varied considerably, and in different states the division between public and private support for the elderly differs. Such differences reflect differences in underlying values, and in the wealth available for such support.

Whilst incapacity is not desirable, many of the other categories of exception permit some agents to fulfill level four needs, and so will be seen as desirable roles. In some cases the role is inherited: for example in some societies bards[11] and, in some cases priests, are hereditary. This can be explained by the need to train the people from a young age, and the need to pass on skills. In other cases, sports being an obvious example, there is a more meritocratic system where payment is linked to perceived ability and in some sports, such as tennis and golf, is quite explicit, with earnings coming from prize money given for success in competitions. In contemporary entertainment in our

---

[11]In other societies those incapable of work may fulfill these roles. Often bards are blind, for example Homer in Ancient Greece and Turlough O'Carolan in eighteenth century Ireland.

society, market forces determine success and failure, and singers perform a less obvious communal function.

Where an exception is only permitted, individuals may voluntarily combine to allow people to not work, expressing the values of a group, and fulfilling further Level 3 needs. For example, in some societies ministers of religion are funded by subscriptions from church members.

5   Example: US Fourth Amendment and the Automobile Exception

We illustrate the use of EAFs to represent norms and exceptions by reference to the US Fourth Amendment. The Fourth Amendment is designed to safeguard the privacy of citizens by guaranteeing:

> "The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."

This amendment has been much discussed in AI since being introduced in [32]. A number of relevant cases are considered in [5]; we will use these cases for our example.

The initial situation can be illustrated by Figure 2a (with Arg2 being *do not search to safeguard privacy rights* and Arg1 being *search to enforce the law* and with P now standing for Privacy and L for Law Enforcement). The matter was felt, however, too important to be left to individual preferences and was resolved by the norm enshrined in the Fourth Amendment. The Fourth Amendment prioritises privacy, and so defeats Arg1. But it also recognises that sometimes searches will be needed, in order to enforce the law in the face of a probable crime, and thus permits searches with a warrant, subject to the safeguard that probable cause has been demonstrated. The situation can be shown as the EAF in Figure 4a.

Suppose, however, that it not possible to obtain a warrant because of the urgency of the situation: for example a car suspected of transporting contraband, which is capable of disappearing into a different jurisdiction before a warrant can be sworn. This was the situation in *Carroll v United States*[12]. In that decision it was held that it was lawful to search an automobile, given probable cause and the impossibility of enforcing the law without the power to search. *Carrol* is the source of what has become known as the *Automobile Exception*. The situation post-Carroll is shown in Figure 4b.

One question is whether the possibility of obtaining a warrant defeats the automobile exception. This was not tested in Carroll, where it was clearly impracticable to obtain a warrant. Some subsequent cases seemed to suggest that police officers thought

---

[12]*Carroll v United States*, 267 U.S. 132 (1925). Other cases discussed are: *Coolidge v. New Hampshire*, 403 U.S. 443 (1971), *Cady v. Dombrowski*, 413 U.S. 433 (1973), *Chambers v. Maroney* 399 U.S. 42 (1970), textitSouth Dakota v. Opperman, 428 U.S. 364 (1976), *United States v Chadwick*, 433 U. S. 1 (1977), *Arkansas v. Sanders*, 442 U.S. 753 (1979) and *California v. Carney.*, 471 US 386 (1985).

if they were dealing with an automobile no warrant was needed, although there were always justices who seemed to think that if a warrant could be obtained, it should be obtained (as Justice Stewart held in *Coolidge v New Hampshire* "The word "automobile" is not a talisman in whose presence the Fourth Amendment fades away and disappears." In *Coolidge* the car was parked in the suspect's driveway and the suspect was in custody. But in other cases the possibility of obtaining a warrant did not seem to matter: searches were held to be lawful when the car was in a commercial garage (*Cady v. Dombrowski*) and even at a police station (*Chambers v. Maroney*). The rationale seems to be that an automobile, when not on private land (as in *Coolidge*), carries lowered expectations of privacy because it is liable to be stopped and inspected under traffic legislation (*South Dakota v. Opperman*).

There were also cases which held it impermissible to search luggage without a warrant, even if the luggage was in an automobile, if the probable cause related only to the luggage and not the automobile as a whole (*United States v Chadwick* and *Arkansas v. Sanders*). The result is the EAF in Figure 4c[13]. Note that the automobile exception does not prevent the police from obtaining a warrant if they choose to do so, it merely frees them to act according to their preferences, We may, however, think that the preferences of the police are such that they are likely to choose to perform a warrantless search if they believe that they are covered by the exception.

In *California v Carney* it was argued that lowered expectations should not apply to a motor home. In the minority opinion Mr Justice Stevens seemed to think that an analogy with luggage might be acceptable.

> "The Court in Chadwick specifically rejected the argument that the warrantless search was 'reasonable' because a footlocker has some of the mobile characteristics that support warrantless searches of automobiles. The Court recognized that 'a person's expectations of privacy in personal luggage are substantially greater than in an automobile,' ... It is perfectly obvious that the citizen has a much greater expectation of privacy concerning the interior of a mobile home than of a piece of luggage such as a footlocker."

But this was rejected by the majority, who held that while in use as a vehicle a motor home had the lowered expectation of privacy associated with vehicles: its primary purpose at the relevant time was not to contain personal effects. Steven's argument about motor homes is therefore not included in Figure 4c.

The advantages of using an EAF are clear. It enables us to see that the 4th Amendment always remains in force, even though there are exceptions to it. Similarly the need for a warrant, if it is possible to obtain one, is not defeated, but can be ignored where there are lowered expectations of privacy. The lowered expectations do not defeat the argument from the possibility of obtaining a warrant, but block its effect. Note, however, that the presence of the automobile on private land does defeat the argument from lowered expectations rather than blocking its effect: expectations of privacy are

---

[13]This situation was modified in California v. Acevedo, 500 U.S. 565, where it was held that the presence of suspicious luggage in a vehicle *ipso facto* gave probable cause to search the whole vehicle, effectively excluding the "luggage" argument.

lowered when the car is in a public space, but not when it is on private land. Similarly the expectations of privacy relating to luggage, which is typically used to transport personal items, remain even when placed in an automobile (at least until *Avecedo*).

6   Concluding Remarks

In this paper we began by exploring Rawls' *Theory of Justice* [31] through the use of computational argumentation, in particular using Extended Argumentation Frameworks [26], with scenarios modelled using Action Based Alternating Transition Systems + Values [2]. We are able to provide the *veil of ignorance* required by Rawls by modelling the scenarios from an individual perspective, which hides the particular roles agents may play, or from the community perspective which hides the roles through aggregation.

We have distinguished between the notions of *promoting values*, as typically used in previous work, and the more fundamental notion of *fulfilling needs*. Whereas values are useful in explaining differences between societies and individual behaviours, needs, which are considered universal to all people, are better able to express why the basic norms of various societies are very similar. This universality is necessary to any Rawls-style reasoning.

We took as our source of needs Maslow's hierarchy [25], in which needs range from those necessary for survival, through those which derive from community membership, to needs relating to self-actualisation. The guiding principle is that preference should be given to lower level needs (such as food) over higher level needs. By applying this preference to competing actions in our scenarios, we can identify the norms endorsed by Rawls' theory.

While this does identify some basic norms, these norms will usually be subject to certain exceptions. What counts as exceptions typically involves a conflict between ways of fulfilling needs at the same level of Maslow's hierarchy. Thus communal belonging at level 3 can be expressed either through joint feasting at carnival, or through listening to bardic renditions of communal stories (or a mixture of both). Similarly different belonging groups may have different objectives: people may come together to pursue mutual interests in arts, sciences or religion, and may choose to support people to engage in these activities rather than in those designed to meet fundamental needs. At level 4 we find needs such as achievement, mastery, and respect from others. Again these may be achieved in a variety of ways, and these ways will differ from society to society. It is here, where the choice of action involves a choice between meeting needs at the same level, that the notion of values becomes relevant: societies may be more or less egalitarian, more or less compassionate, more or less religious and so on.

We distinguish between reasoning to justify norms, which requires reasoning from first principles using needs values and state transitions, from reasoning with norms, in which the arguments justifying the norm are replaced by the norm itself. The focus then moves from competing preferences to the identification of exceptions. These exceptions are themselves justified in terms of preferences, and we find more inter-societal variation than in norms. Thus while most societies will have a norm requiring people to earn their living, the exceptions to such a norm will vary according to the

preferences and ideals of the particular societies. It is to the reasons justifying exceptions that we should look when seeking to identify the values of a society or of groups of citizens.

We also noted that the notion of exceptions helped to make sense of the notion of permissions, which has often proved difficult to understand in normative Multi-Agent Systems approaches such as [1], in which explicit permissions often seem to add little. Here we use them to identify exceptions which the community does not wish to enforce or prohibit, but instead leave it to the voluntary choice of citizens, or particular groups, rather than the citizenry as a whole. The key role of permissive exceptions to norms is to exempt citizens from the obligation to punish certain defined classes of violators, which otherwise is an essential feature of norms if normative collapse is to be avoided.

Finally we should note that some laws relate to very specific features of advanced societies (such as traffic laws). Here it may be better to talk of purposes (as in the origin of this way of resolving conflicts, [13]), reserving values to describe the preference manifested by choosing between different purposes. The chief contributions of this paper are:

- A discussion of how norms can emerge in the manner of Rawls' theory of justice using value-based argumentation, especially where values are interpreted as Maslov's needs and preferences between them justified by reference to levels of Maslov's needs hierarchy;

- The proposal to use EAFs to repreent norms and exceptions, and an illustration of the advantages of so doing.

We believe that EAFs will provide an excellent framework for future representation and discussion of norms.

Appendix: Formal Definitions

Definition 1: AATS [35]. An *Action-based Alternating Transition System* (AATS) is an $(n + 7)$-tuple $S = \langle Q, q_0, Ag, Ac_1, \ldots, Ac_n, \rho, \tau, \Phi, \pi \rangle$, where:

- Q is a finite, non-empty set of *states*;

- $q_0 \in Q$ is the *initial state*;

- $Ag = \{1,\ldots,n\}$ is a finite, non-empty set of *agents*;

- Ac_i is a finite, non-empty set of actions, for each $ag_i \in Ag$ where $Ac_i \cap Ac_j = \varnothing$ for all $ag_i \neq ag_j \in Ag$;

- $\rho : Ac_{ag} \to 2^Q$ is an *action pre-condition function*, which for each action $\alpha \in Ac_{ag}$ defines the set of states $\rho(\alpha)$ from which α may be executed;

- $\tau : Q \times J_{Ag} \to Q$ is a partial *system transition function*, which defines the state $\tau(q, j)$ that would result by the performance of j from state q. This function is partial as not all joint actions are possible in all states;

18

- Φ is a finite, non-empty set of *atomic propositions*; and

- $\pi : Q \to 2^{\Phi}$ is an interpretation function, which gives the set of primitive propositions satisfied in each state: if $p \in \pi(q)$, then this means that the propositional variable p is satisfied (equivalently, true) in state q.

AATSs are concerned with the joint actions of agents Ag. j_{Ag} is the joint action of the set of n agents that make up Ag, and is a tuple $\langle \alpha_1,...,\alpha_n \rangle$, where for each α_j (where $j \leq n$) there is some $ag_i \in Ag$ such that $\alpha_j \in Ac_i$. Moreover, there are no two different actions α_j and $\alpha_{j'}$ in j_{Ag} that belong to the same Ac_i. The set of all joint actions for the set of agents Ag is denoted by J_{Ag}, so $J_{Ag} = \prod_{i \in Ag} Ac_i$. Given an element j of J_{Ag} and an agent $ag_i \in Ag$, ag_i's action in j is denoted by j^i. This definition was extended in [2] to allow the transitions to be labelled with the values they promote.

Definition 2: AATS+V [2]. Given an AATS, an AATS+V is defined by adding two additional elements as follows:

- V is a finite, non-empty set of values.

- $\delta : Q \times Q \times V \to \{+, -, =\}$ is a *valuation function* which defines the status (promoted (+), demoted (–) or neutral (=)) of a value $v_u \in V$ ascribed to the transition between two states: $\delta(q_x, q_y, v_u)$ labels the transition between q_x and q_y with one of $\{+, -, =\}$ with respect to the value $v_u \in V$.

An *Action-based Alternating Transition System with Values*
(AATS+V) is thus defined as a $(n+9)$ tuple $S = \langle Q, q_0, Ag, Ac_1, ..., Ac_n, \rho, \tau, \Phi, \pi, V, \delta \rangle$. The value may be ascribed on the basis of the source and target states, or in virtue of an action in the joint action, where that action has intrinsic value.

The classical notion of a Dung Argumentation Framework (AF) is a tuple $\langle Args, Att \rangle$ where $Args$ is a set of arguments and Att a binary attack relation over $Args$ [16]. The acceptable (justified) sets of arguments are then evaluated as described in [16]. An (audience specific) Value Based Argumentation Framework (aVAF) [4] extends a Dung AF so as to define a partial function val that maps arguments to a value $v \in V$, and an audience constituted in terms of a total ordering $>_a$ over V. Then an attack from argument Y to Z is said to succeed as a *defeat* only if $val(Z) \not>_a val(Y)$. In this way, a Dung AF $\langle Args, Def \rangle$ (with the defeat relation Def replacing Att) is defined and the justified arguments of the aVAF are the justified arguments of $\langle Args, Def \rangle$.

Definition 3: Value Based Argumentation Framework (VAF).
An audience specific Value Based Argumentation Framework (aVAF) is a 5-tuple $\langle Args, Att, V, val, >_a \rangle$ where val is a partial function from $Args$ to a non-empty set of values V, and $>_a$ is a total ordering (audience) on $V \times V$.
Then $Y \in Args$ defeats $Z \in Args$ iff $(Y, Z) \in Att$ and $val(Z) \not>_a val(Y)$.

Extended Argumentation Frameworks (EAFs) [26] extend Dung AFs to accommodate arguments that attack attacks, and so invalidate the success of the attacked attacks

as defeats. For example, if *X* is an argument consisting of reasons in support of the claim that *Z* is a stronger argument (i.e., preferred to) *Y*, then *X* attacks the attack from *Y* to *Z* (denoted $X \twoheadrightarrow (Y \to Z)$). One can thus formalise argumentative reasoning about possibly conflicting preferences and/or value orderings (i.e., audiences). Note that if $X \twoheadrightarrow (Y \to Z)$ and $X' \twoheadrightarrow (Z \to Y)$, then this indicates that *X* and *X'* express contradictory preferences, and so *X* and *X'* attack each other.

Definition 4: Extended Argumentation Framework (EAF). An *Extended Argumentation Framework (EAF)* is a tuple $\langle Args, Att, Att_{Att} \rangle$ such that *Args* is a set of arguments, and:

- $Att \subseteq (Args \times Args)$

- $Att_{Att} \subseteq (Args \times Att)$

- If $(X,(Y,Z))$, $(X',(Z,Y)) \in Att_{Att}$ then (X,X'), $(X',X) \in Att$

We refer the reader to [26] for formal details of how the justified arguments of an EAF are evaluated. Intuitively, EAFs adapt evaluation of the justified arguments in a Dung AF to account for the invalidation of the success of attacks as defeats. Gven a Dung AF, a conflict free set *E* of arguments is said to be acceptable if every argument $Z \in E$ is defended against attacks on *Z*. That is to say, whenever some *Y* attacks *Z*, there is an argument *X* in *E* that attacks *Y* and so defends *Z*. In an EAF, *Z* can also be defended if some *X* in *E* attacks the attack from *Y* to *Z* (i.e., *X* invalidates the success of the attack from *Y* to *Z*).

References

[1] T. Ågotnes, W. van der Hoek, M. Tennenholtz, and M. Wooldridge. Power in normative systems. In *Proceedings of The 8th AAMAS conference*, pages 145–152. IFAAMS, 2009.

[2] K. Atkinson and T. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10-15):855–874, 2007.

[3] K. Atkinson and T. Bench-Capon. Value-based reasoning and the actions of others. In *Proceedings of ECAI*, pages 680–688, 2016.

[4] T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[5] T. Bench-Capon. Relating values in a series of Supreme Court decisions. In *JURIX*, pages 13–22. Citeseer, 2011.

[6] T. Bench-Capon. Transition systems for designing and reasoning about norms. *Artificial Intelligence and Law*, 23(4):345–366, 2015.

[7] T. Bench-Capon. Value-based reasoning and norms. *AI4J–Artificial Intelligence for Justice*, pages 18–21, 2016.

[8] T. Bench-Capon and K. Atkinson. Abstract argumentation and values. In *Argumentation in artificial intelligence*, pages 45–64. Springer, 2009.

[9] T. Bench-Capon, K. Atkinson, and P. McBurney. Using argumentation to model agent decision making in economic experiments. *Autonomous Agents and Multi-Agent Systems*, 25(1):183–208, 2012.

[10] T. Bench-Capon and S. Modgil. Rules are made to be broken. *AI4J–Artificial Intelligence for Justice*, pages 9–17, 2016.

[11] T. Bench-Capon and S. Modgil. When and how to violate norms. In *Proceedings of Jurix 2016*, pages 43–52, 2016.

[12] T. Bench-Capon and S. Modgil. Norms and value based reasoning: justifying compliance and violation. *Artificial Intelligence and Law*, 25(1):29–64, 2017.

[13] D. Berman and C. Hafner. Representing teleological structure in case-based legal reasoning: The missing link. In *Proceedings of the Fourth International Conference on Artificial intelligence and Law*, pages 50–59, 1993.

[14] F. Bex, K. Atkinson, and T. Bench-Capon. Arguments as a new perspective on character motive in stories. *Literary and Linguistic Computing*, 29(4):467–487, 2014.

[15] A. Chorley. *Reasoning with Legal Cases Seen as Theory Construction*. PhD thesis, University of Liverpool, 2007.

[16] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357, 1995.

[17] D. P. Gauthier. *Moral dealing: contract, ethics, and reason*. Ithaca Press, 1990.

[18] F. Guala and L. Mittone. Paradigmatic experiments: the dictator game. *The Journal of Socio-Economics*, 39(5):578–584, 2010.

[19] W. Güth, R. Schmittberger, and B. Schwarze. An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, 3(4):367–388, 1982.

[20] R. Hare. *Freedom and reason*, volume 92. Oxford Paperbacks, 1963.

[21] T. Hobbes. *Leviathan, first published 1651*. Scolar Press, 1969.

[22] A. Jones and M. Sergot. Deontic logic in the representation of law: Towards a methodology. *Artificial Intelligence and Law*, 1(1):45–64, 1992.

[23] M. Lloyd-Kelly, K. Atkinson, and T. Bench-Capon. Emotion as an enabler of co-operation. In *ICAART (2)*, pages 164–169, 2012.

[24] S. Mahmoud, N. Griffiths, J. Keppens, A. Taweel, T. Bench-Capon, and M. Luck. Establishing norms with metanorms in distributed computational systems. *Artificial Intelligence and Law*, 23(4):367–407, 2015.

[25] A. H. Maslow. A theory of human motivation. *Psychological review*, 50(4):370, 1943.

[26] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9):901–934, 2009.

[27] S. Modgil and T. Bench-Capon. Integrating object and meta-level value based argumentation. In *Proceedings of COMMA 2008*, pages 240–251, 2008.

[28] S. Modgil and T. Bench-Capon. Metalevel argumentation. *Journal of Logic and Computation*, pages 959–1003, 2010.

[29] M. J. Osborne and A. Rubinstein. *A course in game theory*. MIT press, 1994.

[30] J. Rawls. Justice as fairness. *The Philosophical Review*, pages 164–194, 1958.

[31] J. Rawls. *A Theory of Justice*. Harvard University Press, 1971.

[32] E. L. Rissland. Dimension-based analysis of hypotheticals from supreme court oral argument. In *Proceedings of the Second ICAIL*, pages 111–120, 1989.

[33] E. Ullmann-Margalit. *The emergence of norms*. OUP Oxford, 2015.

[34] L. Wenar. John Rawls. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2013 edition, 2013.

[35] M. Wooldridge and W. van der Hoek. On obligations and normative ability: Towards a logical analysis of the social contract. *Journal of Applied Logic*, 3:396–420, 2005.
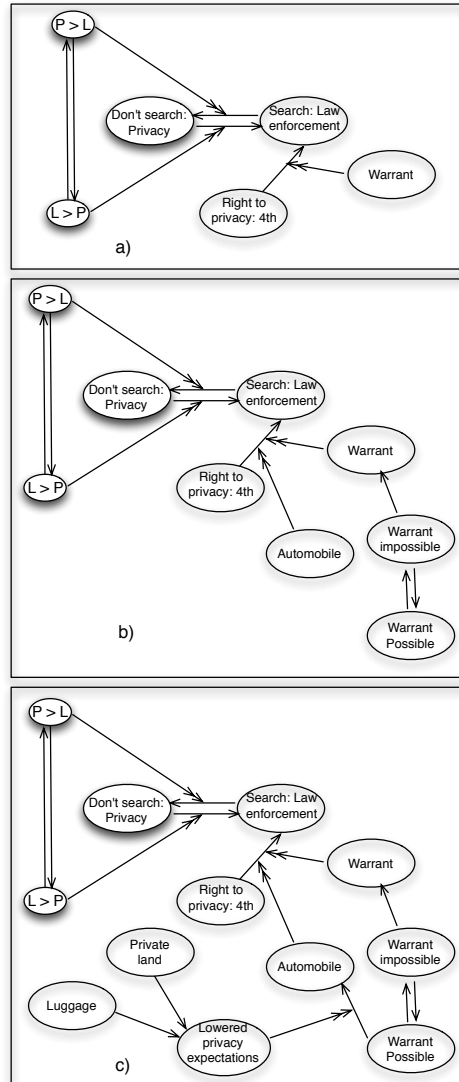
Figure 4: a) Fourth Amendment as EAF. b) After the *Carroll* Case. c) Final Situation of the Automobile Exception.