Getting Up to Speed on Vehicle Intelligence

Leilani H Gilpin, Ben Z Yuan

MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar Street, Cambridge, MA 02139
{ lgilpin, bzy } @ mit.edu

Abstract

Our research explores the development of methodology and supporting technology for combining qualitative and semi-quantitative models with measured data to produce readable explanations of vehicle actions. Control systems in the vehicle should be able to give an accurate unambiguous accounting of the events. These explanations will have to be simple enough for users to understand even when subject to cognitive distractions. For example, when the autonomous system takes over suddenly, the driver will ask why. When an accident happens in a car that is co-driven by a person and a machine, police officials, insurance companies, and the people who are harmed will want to know who or what is accountable for the accident. . In this paper, we present preliminary work towards an explanatory diagnostic system for the vehicle. In particular, we present two models that use measured data to initialize parameters, and then propagate symbolic values to produce concise, understandable explanations of vehicle actions that humans will perceive as a causal chain.

Introduction

Modern automobiles are highly sophisticated, complex, and interconnected electromechanical systems. However, the existing tools for signaling errors, like the "check engine" indicator, only specify the plausible existence of problems without providing any type of explanation or justification. As we move towards semi-autonomous and fully autonomous vehicles, we see a need for cars to be able to produce explanations of their behavior and internal state. These explanations will support the needs of performing maintenance, analyzing driver behavior, and determining accountability for problematic actions. With the addition of sensors, computational elements, and software, pinpointing problems — even simple ones like low tire pressure — becomes more difficult because of the increased range of potential causes. In semi-autonomous driving, understanding the relative contributions of human and machine to an unwanted event, like an car accident, will be important for failure analysis and determination of accountability. Additionally, as incidents occur and provided explanations are challenged by aggrieved parties, these explanations will need to be rigorously justified with appropriate, reliably sourced details.

If we are to trust autonomous agents, ones that may cause fatal injury to humans, then we must first understand that these agents can and will behave in ways other than intended. We must be able to understand why an autonomous agent (mis)behaves in a particular way. Creating the ability for an autonomous agent to provide a coherent explanation of its own behavior is an essential prerequisite to building the necessary trust and confidence in such agents. Our research explores the development of methodology and supporting technology for combining qualitative and semi-quantitative models with measured data to produce concise, understandable symbolic explanations of actions that are taken by a complex system that is built out of many components.

Previous Work

There have been many contributions in the space of reasoning for qualitative change. The term incremental qualitative (IQ) analysis was coined in Johan De Kleer's PhD thesis (Kleer 1979). Incremental qualitatives represent quantities by how they change when a system itself is changing. Incremental qualitatives are represented by four values: increasing, decreasing, no change, unknown change. In his thesis, DeKleer describes this qualitative algebra, which we extend in this work to represent different algebraic operations. De Kleer and Forbus also wrote a book on efficient logic truth maintenance systems (LTMSs) (Forbus and Kleer 2003) to propagate constraints, however, it lacks the verbose explanatory capability necessary for our application.

Using back propagation to reason about complex behavior has also been well studied in the field. Models that can merge measurements with deductive reasoning and that can track dependencies are captured with propagators (Radul and Sussman 2009). This language is a way to build distributed, concurrent computational models, based on the idea that computational elements, called propagators, are autonomous machines interconnected by shared cells. Each propagator continuously examines its neighbor cells adding information to some, based on deductions it can make from the information in others. Our models are implemented in MIT/GNU scheme with the propagator framework.

In vehicle specific modeling, Stuss and Fracci were able to present a qualitative model a vehicle braking system (Struss and Fracci 2014). Their major contribution is a

model-based automation of failure-modes-and-effects analysis (FMEA), with the specific application to a vehicle braking system. While their results do not include explanations, their model was motivating in our own research. They are able to infer braking component behavior from the models' inputs.

As far as we know, this is the first application of qualitative reasoning to provide explanations of complex and comprehensive vehicle actions.

Methods

We have developed a two-step process to provide explanations from vehicle logs. The first step detects what events happened. We constructed a set of rules to recognize events (not all encompassing), and detect if these events occurred during a vehicle log. The second step finds reasons for those events.

In the first step, we take a CAN bus log as input to analyze what events happened during a specified time interval. This analysis includes smoothing noisy data, performing edge detection to find out when particular events occurred (e.g. when did the operator apply the brakes), and interval analysis to see how particular intervals relate to each other (e.g. did the car slow after the brakes were applied?). Using this analysis, we were able to construct a story of what happened in a particular car trip and detect particular events of interest (e.g abrupt changes in speed and braking, and dangerous maneuvers like skids).

In the second step, we take a particular interval of interest and backtrack through the model to construct an explanation. We developed two different models to explain vehicle mechanical processes and physics in a human readable form. We constructed a qualitative model of the vehicles' mechanical components, which represents and explains the dependencies between the vehicles' mechanical components. This model is purely qualitative. The vehicle actions are represented symbolically, in qualitative terms like increasing, decreasing, no change, and unknown change. While this model is easy for humans to understand, it lacks the level of detail needed to explain more sophisticated vehicle actions like skids. So we have also developed a semi-quantitative model of car physics. This model infers the overall effect on the normal forces and frictional forces on the wheels from the reported lateral and longitudinal acceleration during a particular interval. Then, these effects and their consequences are explained qualitatively to the user.

Vehicle Data

Collecting the necessary data from an actual vehicle, especially data corresponding to an accident scenario, is a challenging prospect. We instead produced the necessary information from a plausible vehicle simulation. We developed a basic vehicle simulation using the Unity game engine, including physical properties like tire friction and basic vehicle internals, at a level of fidelity sufficient to test our analysis pipeline. This simulation, in response to user control of a simulated vehicle, produces data traces corresponding to quantities that are tracked by commonly installed vehicle sensors communicating on a typical vehicle CAN bus,

like accelerometer data, wheel rotation rates, and driver input state. To generate the necessary data corresponding to an accident scenario, we built a model highway interchange in the simulation, and then drove the simulated vehicle in a variety of accident regimes.

We considered as a motivating scenario an example of a car entering an oversteering skid on a freeway offramp. This situation can result when brakes are applied during an otherwise controlled turn when the vehicle is moving at high velocity, causing loss of friction on the rear wheels resulting in lateral motion of the rear of the vehicle. We did several simulation runs to replicate the described event, gathering test traces to use for the developed analysis pipeline.

We have developed a series of steps to find "intervals of interest" in order to analyze the most prevalent and interesting events. These intervals, represented by a starting time and ending time, designate a significant change in the vehicle's position, one or more of the mechanical components, or physical state. The process of finding these intervals begins by smoothing the noisy data using an average pass filter. We then identify specific intervals where edge events occurred. For example, we find a set of intervals when the brakes were applied, or when the car was accelerated. We filter our data to find intervals that adhere to these edge detection rules, resulting in braking intervals, accelerating intervals, right turning intervals, etc. We then filter these edge event intervals that satisfy the temporal relationship description of particular events. For example, once we have braking intervals, we want to find the intervals within that where the vehicle is turning, or where the vehicle came to a complete stop. We represent these events as a combination of intervals using Allen's Interval Algebra (Allen 1983) shown in Table 1 to find intervals that work together to make an "interval of interest". We then use these intervals of interest to form a story of what happened during a particular time span.

X before Y	XXX YYY
X equal Y	XXX YYY
X meets Y	XXXYYY
X overlaps Y	XXX YYY
X during Y	XXX YYYYY
X starts Y	XXX YYYYY
X finishes Y	XXX YYYYY

Table 1: This is a complete set of the possible relationships among time intervals, as determined by James F. Allen in 1983.

Being able to name these relationships is the first step toward symbolic reasoning and explanations involving time. We then use these relationships to create a summary of events, like the one shown in Table 2. Now that we have a set of interesting observations, we needed models to explain in greater detail the processes underpinning each event. Since

Time	Event
18:10:25.333	GPS: Heading 321.16, Speed 60.3mph
18:10:26.500	Operator: Brake 0.35, Steer 5.0
18:10:26.560	Driver assist: Brake 0.45 :-)!
18:10:27.867	GPS: Heading 353.84, Speed 52.1 mph
18:10:29.970	Operator: Brake 0.90, Steer 9.3
18:10:30.010	Wheel Rate Monitor: Skid
18:10:30.040	GPS: Heading 28.27, Speed 0.0mph
18:10:30.070	Wheel Rate Monitor: Skid
18:10:30.170	Operator: Brake 0.91, Steer 6.6
18:10:32.933	GPS: Heading 129.08, Speed 0.2mph
18:10:35.140	Operator: Brake 0.93, Steer 0.0
18:10:35.467	GPS: Heading 121.52, Speed 0.0mph
18:10:38.670	Stopped

Table 2: Summary of the "intervals of interest" accumulated during a over-steering skid

the mechanical and physics phenomena in the vehicle are quite complicated, we rely on a set of qualitative algebras to be able to explain the events in more readable language.

Qualitative Algebras

We define a set of qualitative algebras to explain our vehicle scenarios. Our first and most used algebra, the qualitative increment, was first coined in De Kleer's PhD thesis (Kleer 1979). However, in our vehicle models, we are not only concerned with qualitative increments like increasing, decreasing, etc. Since mechanical components have different actions not described by incremental values (like tightening and loosening), and vector forces on the vehicles' wheels have direction and magnitude, we developed a set of qualitative algebras to represent the descriptions of mechanical components and force vectors. We defined four qualitative algebras: a qualitative increment, action, direction, and magnitude.

The qualitative increment algebra is essentially a qualitative description of the first derivative. In our current work, these qualitative descriptions of the first derivative are sufficient to explain the vehicle phenomena. However, for future work, we may need to use more precise quantitative descriptions. In anticipation of that, we have developed a system to keep track of derivative values on the cells in the art of the propagator system.

The second qualitative algebra, the qualitative action, describes mechanical changes within an interval as a set of four actions: tightening, loosening, no action, and unknown action. It may appear as if the qualitative action is unnecessary, since it has a surjective mapping to the qualitative increment, where tightening is a qualitative increment value of increasing, loosening is a qualitative increment value of decreasing, no action maps to no change, and unknown action maps to unknown change. However, having the qualitative action makes our explanations much easier to report: while the description of "tightening a caliper" is easy to understand, "increasing a caliper" is a bit ambiguous.

The third qualitative algebra, the qualitative direction, has two values: an incremental change description and a direction description. The incremental change is defined as the

qualitative increment. The second value, the direction description, describes a lateral direction in four values: left, right, neutral, and unknown direction. This direction is with respect to a point of reference. In our research, the point of reference is usually the center of mass of the vehicle.

The final qualitative algebra, a qualitative position also has two values: a lateral description and a longitudinal description. The lateral description is defined with respect to the direction description defined in the qualitative direction. The longitudinal direction is defined with respect to four values: front, back, neutral and unknown.

After defining the algebras, it is necessary to define the combinations of these values so that the algebraic descriptions can work together to provide a comprehensive explanation. It then becomes important to define the resulting qualitative type from different combinations, so if two qualitative values are added, the resulting qualitative type value makes sense. For example, if a qualitative increment and a qualitative action are added together, the resulting value with be a qualitative increment, because the qualitative action can be directly mapped to the qualitative increment. Since there are only one type of combinations of different qualitative algebras in our model so far: combining qualitative increments and qualitative actions, we leave the definition and justification of a qualitative hierarchy to future work.

Model

We constructed two models, which combined with generated data, provide human-readable explanations of vehicle behavior. The fully qualitative mechanical model has rules which describes the relationships between mechanical components of the vehicle. For example, when the tire brake pads are engaged, what other mechanical components are affected? The mechanical model also provides basic reasons for causality. Using the same braking example, when the tire brake pads are engaged, the wheel rotation rate is certainly decreasing. However, if the wheel rotation rate is decreasing, that does not necessarily imply that the tire brake pads were engaged. In fact, the vehicle could be going uphill, or the vehicle could slowing down due to the force of friction on the tires. In order to provide more detailed explanations of why a particular event occurred, we developed a semi-quantitative physics-based model which quantitatively calculates forces on the vehicles wheels and combines that information with measured data to provide qualitative explanations. Both models are written in MIT/GNU Scheme and rely on the existing propagator system.

Propagator System The Art of the Propagator System is a qualitative reasoning system that maintains dependencies that can be used to construct an explanation of how a command to the effectors or an intermediate value was determined. Part of those dependencies are causal chains that come from the expectations determined by the model, and some will be constraints coming from recorded data. The propagator system, described in (Radul and Sussman 2009) can merge measurements with deductive reasoning and track dependencies. Propagators implement a way to build distributed, concurrent computational models interconnected

by shared cells. Each propagator continuously examines its neighbor cells adding information to some, based on deductions it can make from the information in others.

Consider, for example, the block diagram of the throttle-control system in Figure 1. A qualitative model of this system can be made from this diagram. Each wire can be modeled by a propagator cell that holds information about the signals on that line. Each block in the diagram can be modeled by a propagator that constrains information contained in the cells that model the wires incident on it. Although the diagram indicates that information flows in specific directions, propagators can make deductions about inputs from information about outputs as well as in the indicated direction.

In any particular situation the models and the data will converge to some approximate agreement if the models are good enough. There will be many points of agreement where the data will be what is expected and there may be points of disagreement. Points of disagreement indicate inadequacies of the models or failures of some part of the real mechanism to do what is expected of it.

Qualitative Mechanical Model The qualitative mechanical model describes the interactions among the vehicle's mechanical parts. The mechanical components are modeled from the diagrams and descriptions in an automotive handbook (Reif and Dietsche 2014). Currently, the mechanical model is relatively simple, only modeling the braking system, steering system, and engine as shown in Figure 1. The sensor output is not yet implemented in our system. One reason for the over-simplicity of the model is that is our simulated data does not represent all the vehicle data. But, the more pressing reason for a model simplicity is that we expect qualitative values. Therefore, the rules of the system are intuitive, and the system does not require complex equations.

This mechanical model is initialized by a specific time interval, where each data value (brake pad pressure change, steering change) is a qualitative increment, representing the qualitative change during that time span. Then, the model is run by propagating the data-initialized qualitative values through the system. Specific model values can be queried using `explain-readable` and `inquire-readable`, which are explained in detail in the Readable Explanation section.

Semi-quantitative Physics Model The physics model calculates the various forces on the vehicle's wheels, and then constructs causal chains; some coming from expectations determined by the model, and some coming from recorded data. Unlike the mechanical model, which is initialized by recorded data, and then queried for values and explanations, the physics model is initialized by recorded data, but also uses recorded data values to provide evidence in its deductions. The forces that are calculated and then explained are shown in Figure 2 from the lateral and top down view. These forces and the figure are not meant to be inclusive of all the forces on the vehicle.

The physics model is especially important when the vehicle's actions are unexpected and not interesting from the
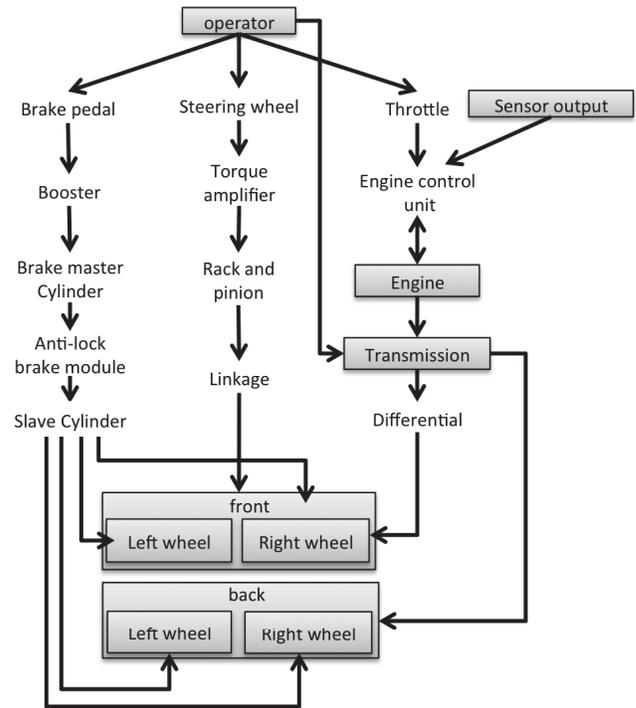

Figure 1: A diagram of the mechanical model that describes the interactions between the vehicle's mechanical parts.
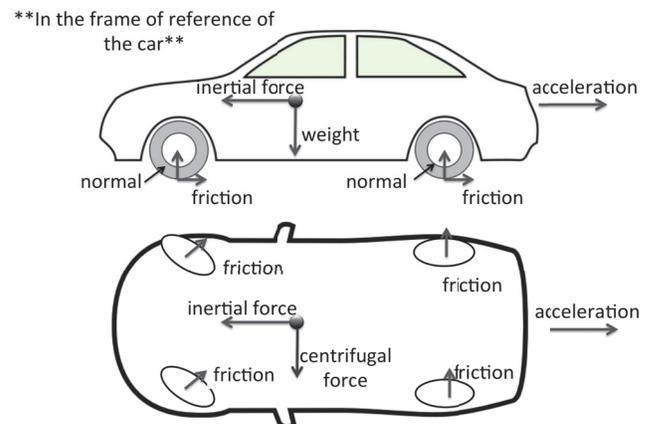

Figure 2: A visual guide to the car forces that are calculated and used in the semi-quantitative physics model.

mechanical model standpoint. For example, if a skid occurs, the mechanical model will only be able to describe the wheel rotation rates in terms of qualitative increments, which does not provide the explanation for a skid. Instead, using the physics model, a skid can be identified by the rear or front wheels losing traction. The wheels lose traction by a decrease in normal force.

It is important to note the distinction between deductions

and expectations. In our work, we deduce based on a casual chain of expectations. Our assumptions are certain; we expect the mechanical devices within the car to act a certain way, and at the same time, we can expect the physics of the vehicle to act a certain way.

The major contribution of our work is designing a system for combining observations with data and qualitative physics explanations.

Readable Explanations   We have made a first pass at providing readable explanations from the dependency tracking elements of the propagator system. Most of our explanations are rule-based, or a somewhat more readable version of an `explain` command, which is implemented in the propagator system. The propagator has two explanatory functions `inquire` and `explain`. The `inquire` command displays the immediate values of the inquired cell, and `explain` command recursively displays every dependency in that cell's inquiry in the system. We implemented two other explanatory functions, `inquire-readable` and `explain-readable`, which display the dependencies tracked through the system in more readable language. For example, the `inquire` and `explain` display the cell's dependent functions and values without explaining them, whereas `inquire-readable` and `explain-readable` explicitly cites the reasons, inputs, outputs, premises, and displays the cell value in readable form.

Results

We have implemented two models, by compiling the rules to propagator structure that describe the interactions among parts. The experiments displayed in this section are not inclusive of all our findings, however, we chose to display some of the more "interesting" events.

Examples from the Mechanical Model

The mechanical model propagates observed data through the different mechanical subsystems of the vehicle. The model is initialized by the observed data, and then those values are propagated through the system appropriately. Then, the user is able to query specific components of the system, and a readable explanation can be displayed through the appropriate commands below:

```
>> (inquire-readable
(hydraulic antilock-brakes))

(hydraulic antilock-brakes) has the value
increasing change
  Reason: function (+)
    inputs: front-brakes-pressure
              (booster antilock-brakes)
    outputs: (hydraulic antilock-brakes)
  Premises:
  (brake-change-from-initialized-interval)
>> (inquire-readable left-front-wheel)
left-front-wheel has the value
decreasing change
  Premises:
  (wheel-change-from-initialized-interval)
```

From the model, we are able to query different internal mechanical devices, like the hydraulics, and we can also query the data directly, like the left-front-wheel rotation rate. Notice that the left-front-wheel has no inputs because it is a data input. The initialization, which we named to be `wheel-change-from-initialized-interval`, is the premise which tells the `left-front-wheel` to be exactly the value decreasing change.

While this model is useful for debugging mechanical systems and, it takes the physics knowledge to be able to model complex vehicle behavior like skids.

Examples from the Physics Model

The physics model propagates the underlying forces on the wheel of an average front-wheel drive sedan. The model is initialized by the acceleration forces (both lateral and longitudinal) and calculates what is the appropriate magnitude of forces on the car's wheels. The first example is of skid behavior.

```
(explain-readable-forces normal-forces)
REASON: rear-wheels-force decreased AND
its magnitude exceeds the traction
    threshold.
    Since the rear wheels lost traction
    the friction of the contact patches
    MUST HAVE decreased;
    so, the normal forces MUST HAVE
    decreased.
  Consistent with the accelerometers.
QUALITATIVE TIRE SUMMARY:
    The left front normal force decreased.
    The right front normal force increased.
    The left back normal force decreased.
    The right back normal force decreased.
```

The above explanation was automatically generated by explaining the normal forces during an oversteering skid interval. The more detailed explanations are generated by aggregating quantities and qualitative reasons that are set after running the physics model.

```
(explain-readable-forces normal-forces)
REASON:  right-wheels-forced increased
its magnitude is within traction threshold.
    Since the right wheels are gaining
    traction
    the friction of the contact patches
    MUST HAVE increased.
    so the normal forces MUST HAVE
    increased
    So the car is turning left safely.
  Consistent with the steering
  and accelerometers.
QUALITATIVE TIRE SUMMARY:
    The left front normal force decreased.
    The right front normal force increased.
    The left back normal force decreased.
    The right back normal force increased.
```

The above explanation was automatically generated by explaining the normal forces during a left turn.

Future Work

The models we propose in this paper are a first step towards developing an explanatory system for the vehicle. Our immediate future work is to combine the physics and mechanical models. With a combined model, and more detailed sensor data, we could model more complex physics over complex terrain. For example, for a turning radius and a friction coefficient for what speed will the car skid? And if we observe that the car is decelerating, is this caused by the brakes, or by the hill or both? Preliminarily, we have developed a naive query language to find edge cases, but now we will need a more sophisticated language, and possibly multiple languages for different parts. We will also advance the language for accepting queries and building human-understandable explanations. In future work, we hope to be able to use real CAN bus data, and we will apply our combined system to real, existing logs and report on the results.

We will investigate what mechanisms are necessary to provide dynamic explanations in a running vehicle. This may require some optimization in order to be used in real time, but the more important result would be ways that qualitative deductions from the running logs might be used to improve the control strategies in vehicles with significant autonomy.

When we model a system like a modern vehicle, we are really creating a model of the "mind" of a car. A "modern" vehicle is a system made up of lots of pieces that individually handle different tasks, like anti-lock braking, power steering, and obstacle detection. Similarly, the human mind can be thought of as a set of various pieces that do different jobs and speak very different internal languages. Our goal is to construct a model of a car mind made of up different parts that can "share stories" with each other. We have explored this idea in our work so far in a limited setting with cells in a propagator-based system, where the "sharing" of information is done via the propagator rules. We take this a step further by applying these ideas to full-sized systems, in which the communicating parts are entire electromechanical modules, and the stories shared with each other collectively contain enough detail to permit reconstruction of adequate explanations of particular phenomena.

Discussion

We are building the underlying technology and strategies to model layered systems of communicating agents. In such a system, each component is represented by an agent that has a specific task or jobs. Agents need to be able to talk to and understand their neighboring agents in an appropriate language. Each agent needs to be able to tell a plausible story about what is going on around it. In the case of the car example, the braking components should be able to tell compelling stories about what the tire sensors and steering components are doing. If the brakes are not able to develop a plausible story, then that is evidence that something is wrong in the system. However, every agent is not fully aware of the behavior of all agents in the system: agents that are levels apart are not able to directly communicate. For example, it does not make sense that the high level route planning agent can directly make stories about the actuation components like the braking agents, since, these agents are several levels apart. Our initial vehicle model agent will have levels similar to that of our mechanical model.

Explanation is relevant here because every agent is constantly developing a plausible story about its neighbors. Without explanations, how do we know if the autonomous vehicles are working in our best interest? When something goes wrong, as it inevitably will, we are able to "listen in" on this system of communicating agents to find implausible stories. Those stories can be used as evidence that something is wrong. This system can also find and assign blame, if that is necessary, by backtracking through the stories provided by agents and pinpointing ones that do not make sense. And in general, this system could be used to debug systems involving machines and people. The human operator itself is an agent in the system. If the driver acts irrationally, then the steering and braking agents will have unsatisfactory explanations. Similarly, if the steering and braking agents are not working as intending, the human operator will develop an unsatisfactory explanation for these neighboring agents.

We have seen the deployment of autonomous agents that cannot explain themselves. Even the best problem solvers, like AlphaGo and Deep Blue, cannot explain their actions, whether they are right or wrong. These machines must be able to use their records to tell a coherent story about the reasons for their decisions or actions. And further, that story must be understandable by other agents, including humans, and it must be able to be challenged in some way. If it is determined that the explanation provided is an inadequate or inappropriate justification for the decision or action, then the agent should be able to be corrected. It should be possible to modify the behavior of the agent, so that no similar explanation can be used to justify a similar action in the future.

Only when systems are auditable and articulate enough to explain their actions and decisions as a symbolic story will we be confident of their competence.

References

Allen, J. F. 1983. Maintaining knowledge about temporal intervals.

Forbus, K. D., and Kleer, J. D. 2003. *Building Problem Solvers*. MIT Press.

Kleer, J. D. 1979. Casual and telelogical reasoning in circuit recognition. *AI Technical Report*.

Radul, A., and Sussman, G. J. 2009. The art of the propagator. *CSAIL Technical Report*.

Reif, K., and Dietsche, K.-H. 2014. *Automotive Handbook*. Bosch, 9 edition.

Struss, P., and Fracci, A. 2014. Modeling hydraulic components for automated fmea of a braking system. In *Proceedings of the Annual Conference of the Prognostics and Health Management Society 2014*.