Auditability and Traceability

Are these requirements for Autonomous (Weapon) Systems?

Abstract: Remotely piloted aircraft are everywhere: the Predator (General Atomics), Global Hawk (Northrup Grumman), ……(Lockheed), …..(Boeing), ……(DIY Drones). They are the new weapons of war and they allow the owners to keep their warfighters out of harm's way. But each one requires a team to execute a mission. It is clear that operational manpower requirements will drive the transition initially from a team to support one system, and then toward an individual that can command a swarm of systems. To reach this level of capability, these systems will have to migrate from human-in-the-loop, to human-on-the-loop, and eventually to human-outside-the-loop where full, or almost-full, autonomy will be necessary. This paper opens the discussion of whether these systems need to be auditable. It separates auditability from traceability. It discusses whether the policies need to be traceable to the authors / owners of the policies executed by the autonomous systems so they can be held responsible.

This paper decomposes the potential requirement for Auditability and Traceability in Autonomous Systems into a number of areas. It attempts to identify pros and cons so that global policy makers can make the best decision possible.

It exposes the following questions for comments:

- Why might auditability be necessary? (If not, why not)
- Why might traceability be necessary? (If not, why not)
- Who might be the audience for auditability/traceability results (and why)?
- What might one want to gain from auditing the behavior of a system? (What information)
- How might one determine if the auditing process has been successful? (or can be successful)
- How might one determine the cost of auditing the behavior of a system? (Recurring cost and incremental cost in order to determine what is acceptable)
- What system considerations need to be accounted for when constructing an auditable system? (Different approaches / different levels of auditability)
- How might one expect to audit a system? (Different approaches / different levels of auditability)

This document uses a new approach for exposing and discussing the topic. The topic was first raised on the InnovationHub sponsored by NATO Allied Command Transformation at (http://innovationhub-act.org/drupa/node/598). NOTE: a username and password are required for access.

In this case a work-in-progress document is managed by Compsim. It has collected information on Auditability and Traceability from the InnovationHub. That information has been extended with local brainstorming. Using Compsim Management Tools (a knowledge capture and decision-making tool) interactive discussion points were created that are included in this document. An introduction into how to navigate this interactive document is provided in the Appendix.

Each sub-issue is presented as an interactive information map. That is followed by a summary (all-inclusive) information map. To review the information, one should review the procedure defined in the Appendix.

Why might auditability be necessary? (If not, why not)

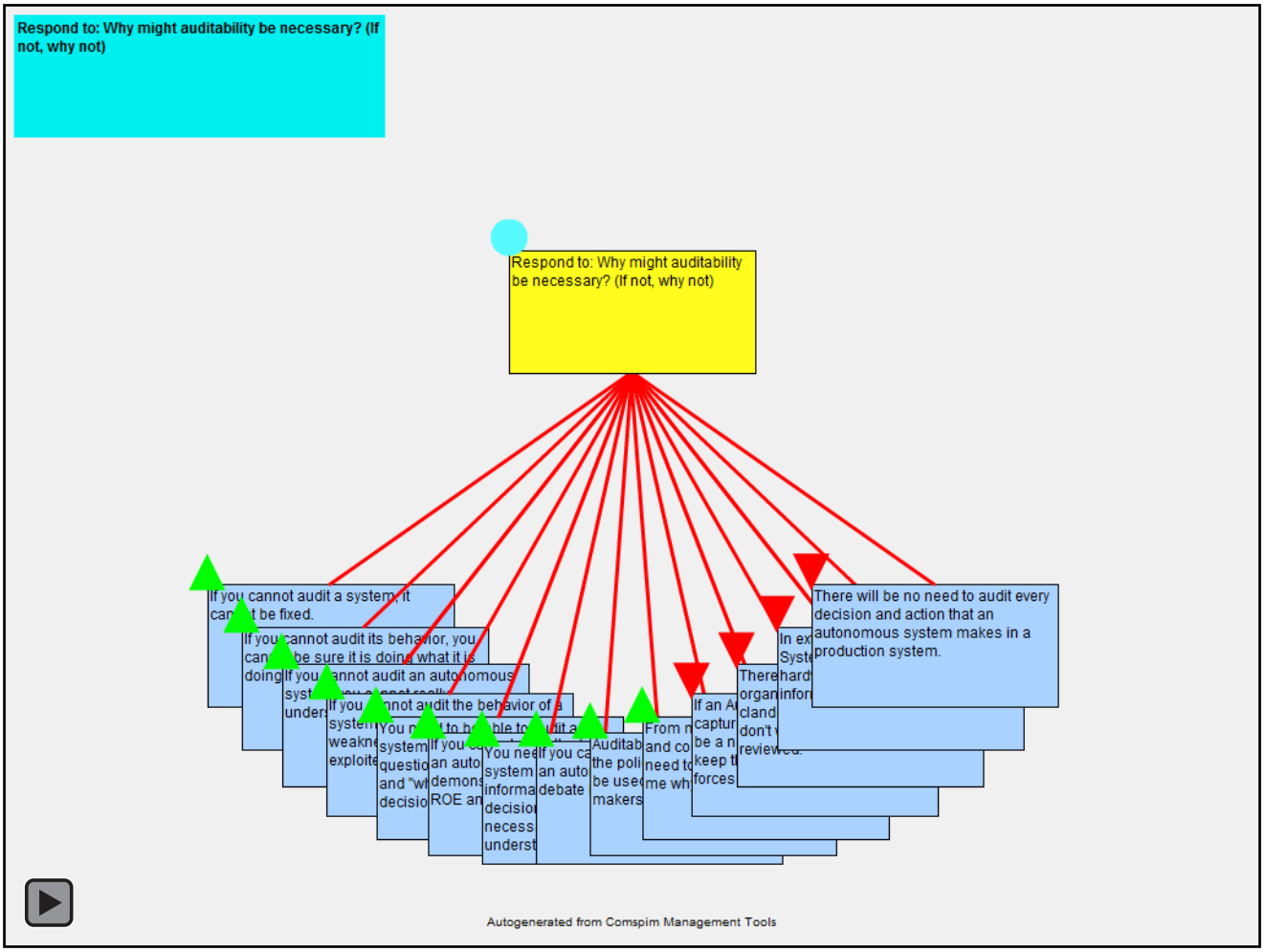

Figure 1

Why might traceability be necessary? (If not, why not)

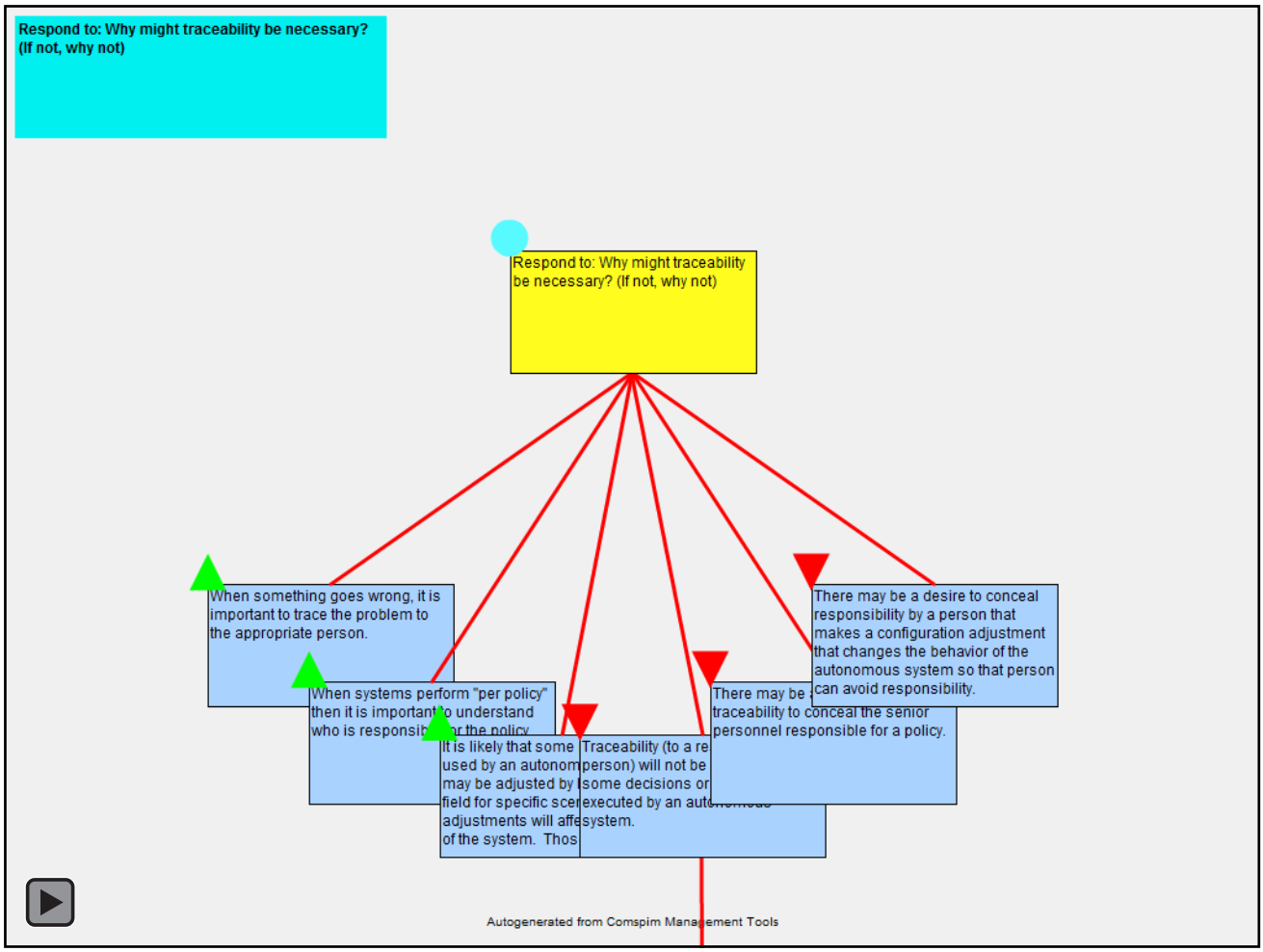

Figure 2

- **Who might be the audience for auditability/traceability results (and why)?**

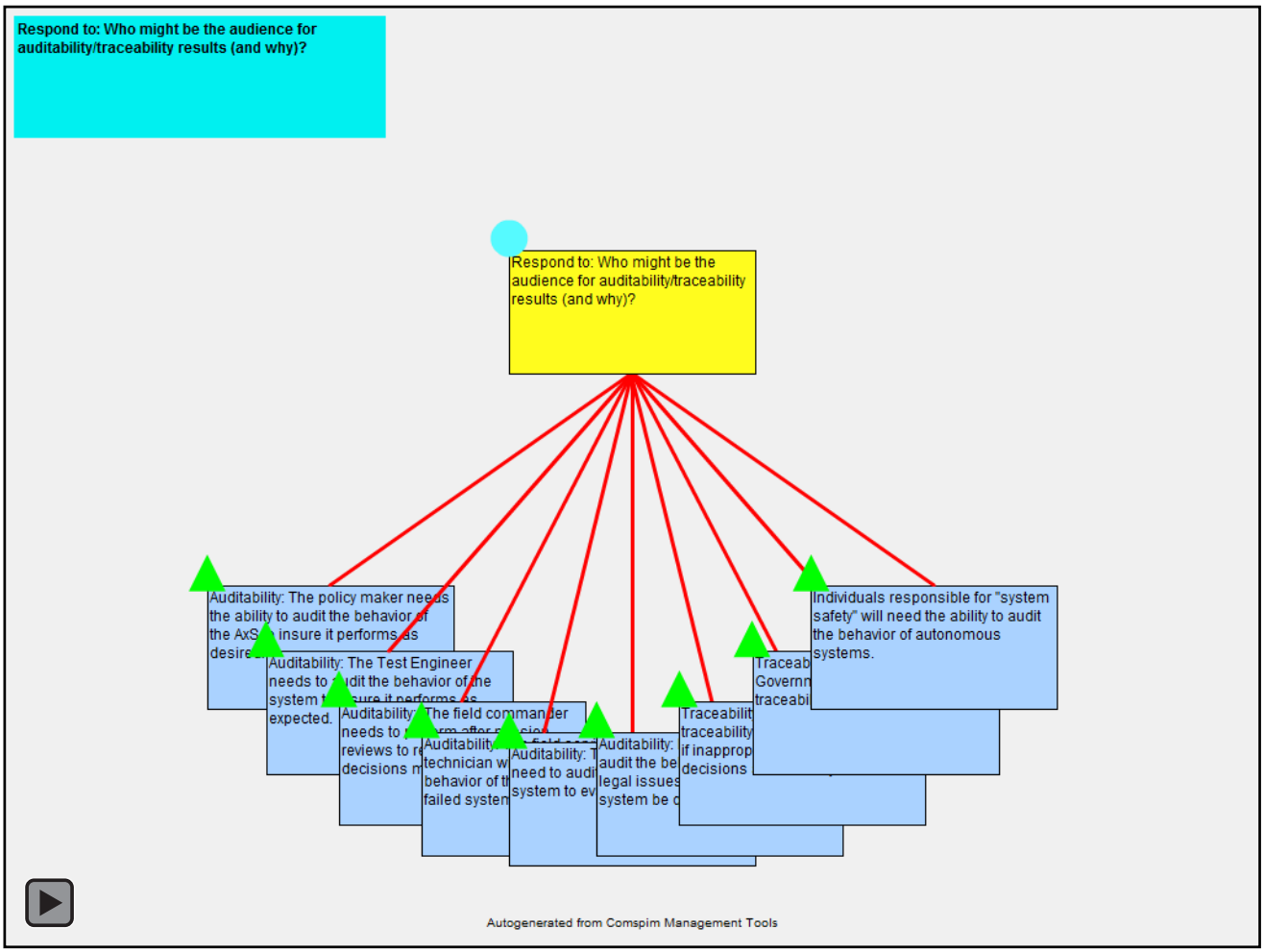

Figure 3

What might one want to gain from auditing the behavior of a system?

(What information)

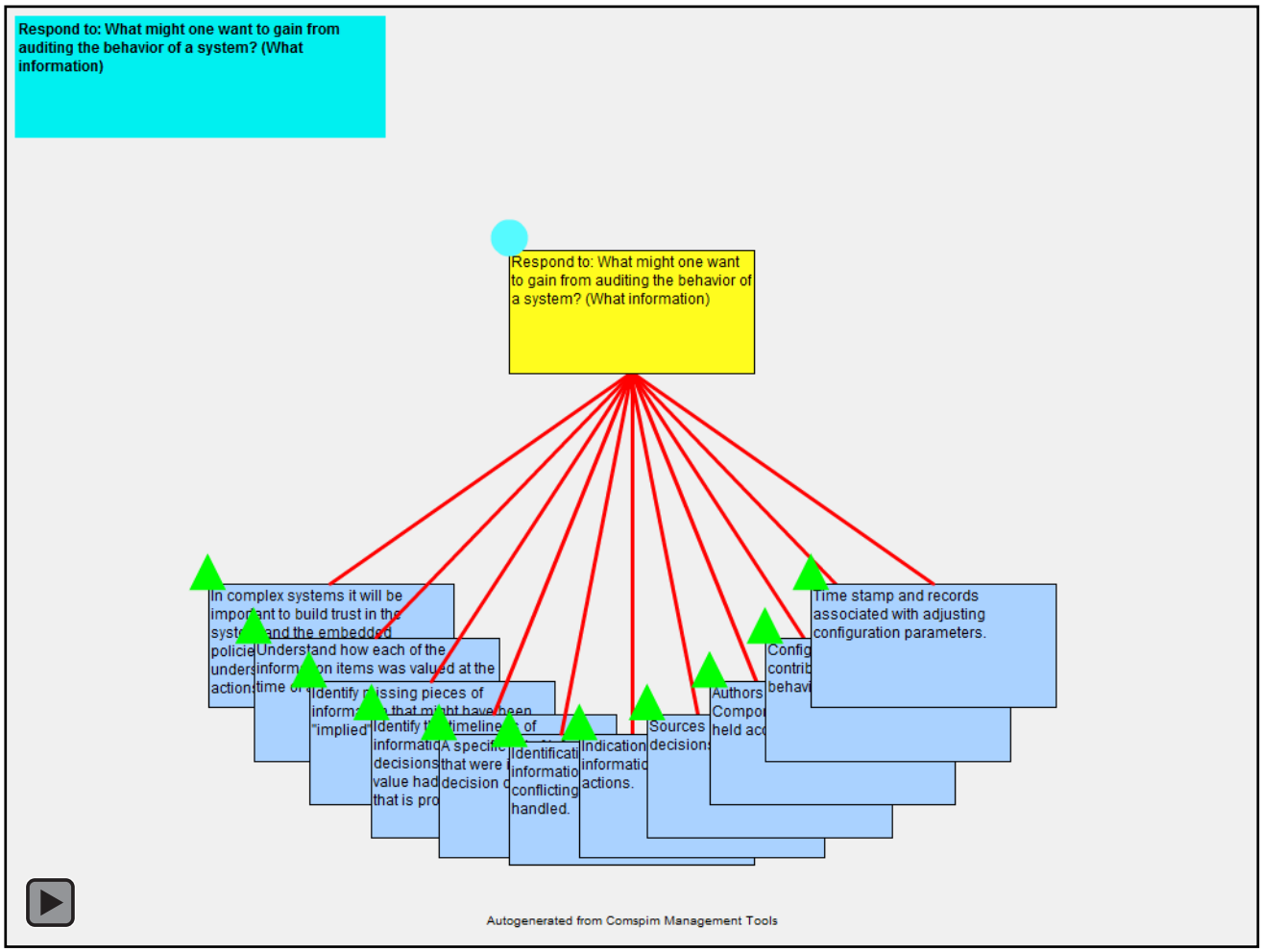

Figure 4

How might one determine if the auditing process has been successful?

(or can be successful)

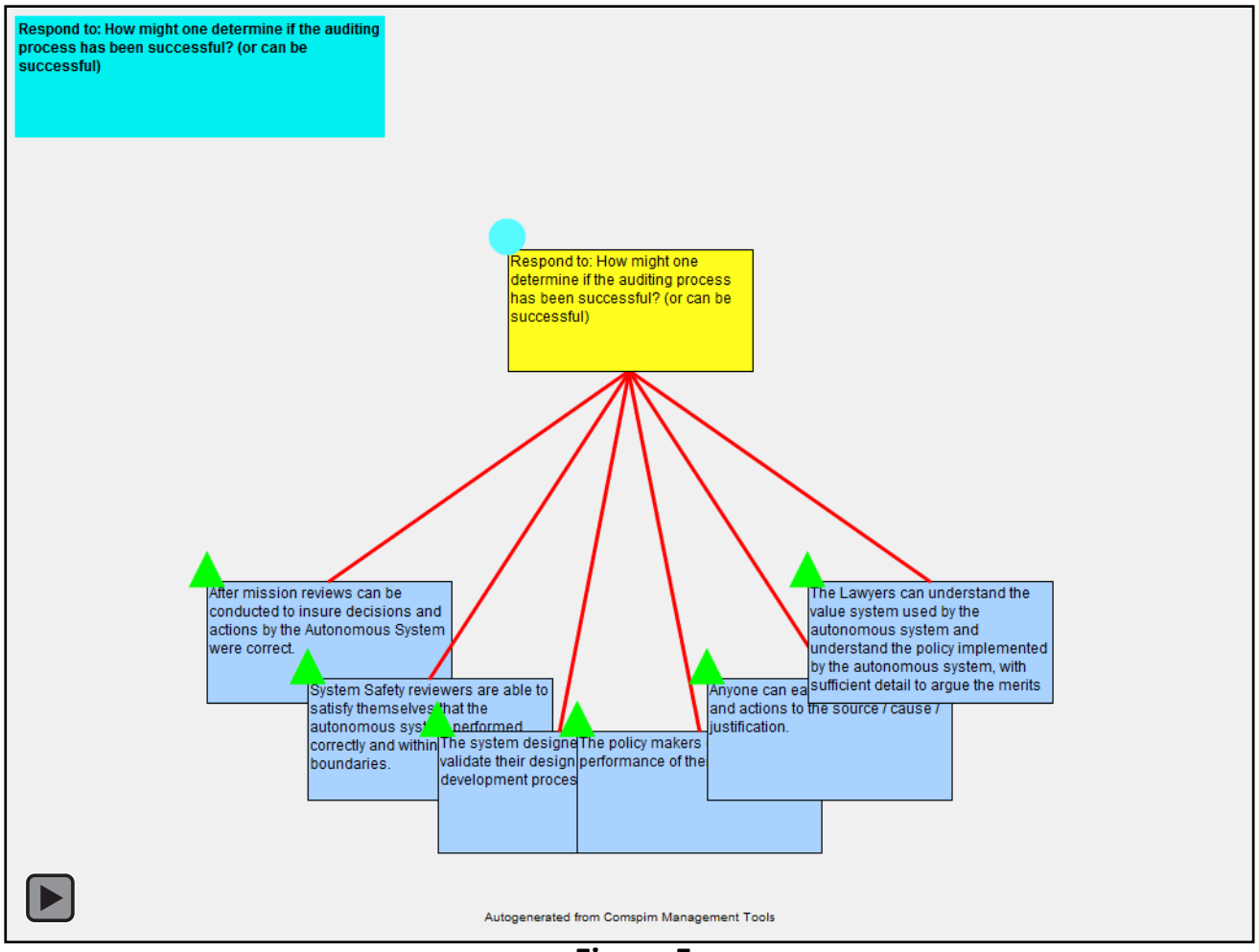

Figure 5

How might one determine the cost of auditing the behavior of a system? (Recurring cost and incremental cost in order to determine what is acceptable)

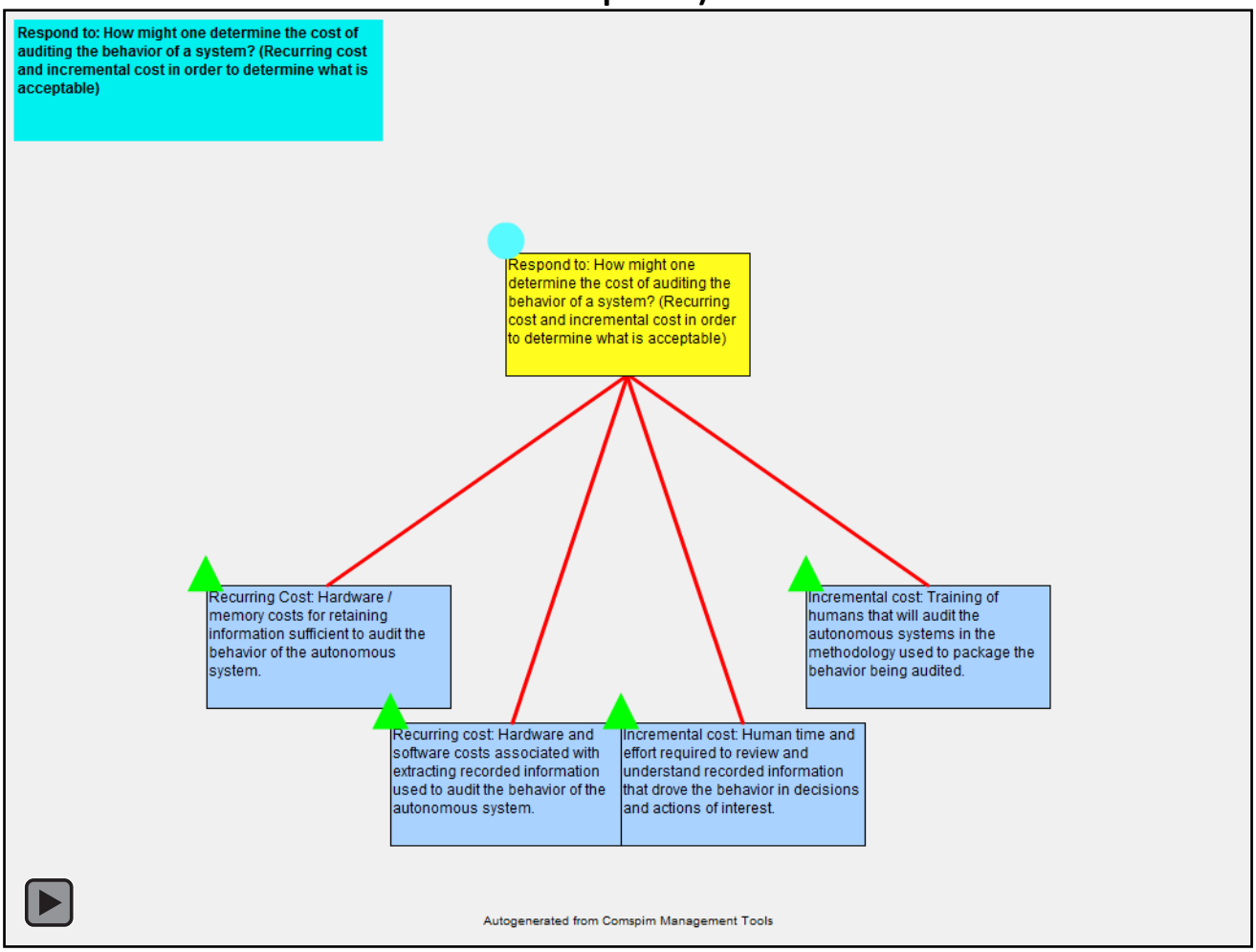

Figure 6

What system considerations need to be accounted for when constructing an auditable system?

(Different approaches / different levels of auditability)

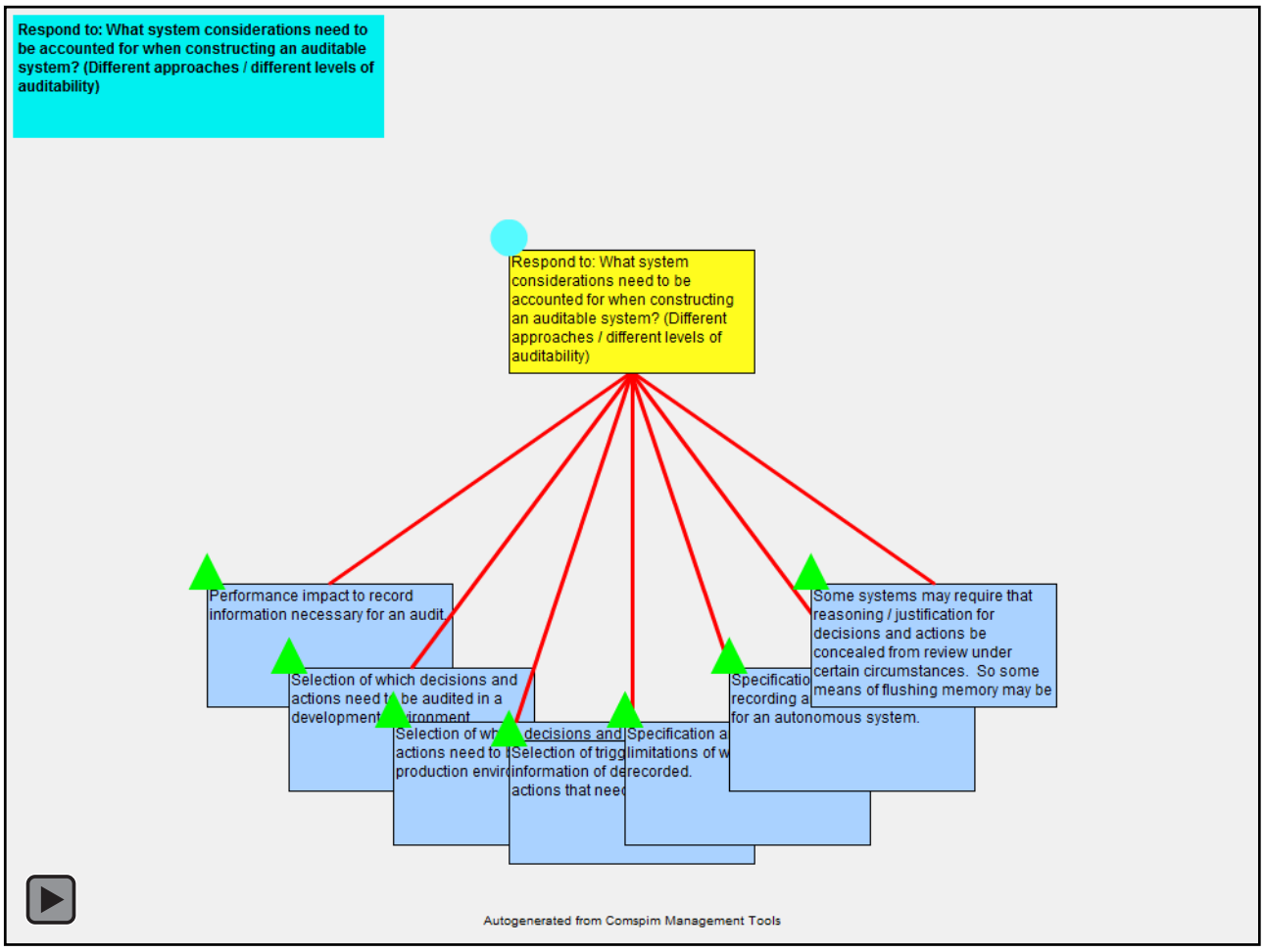

Figure 7

How might one expect to audit a system?

(Different approaches / different levels of auditability)

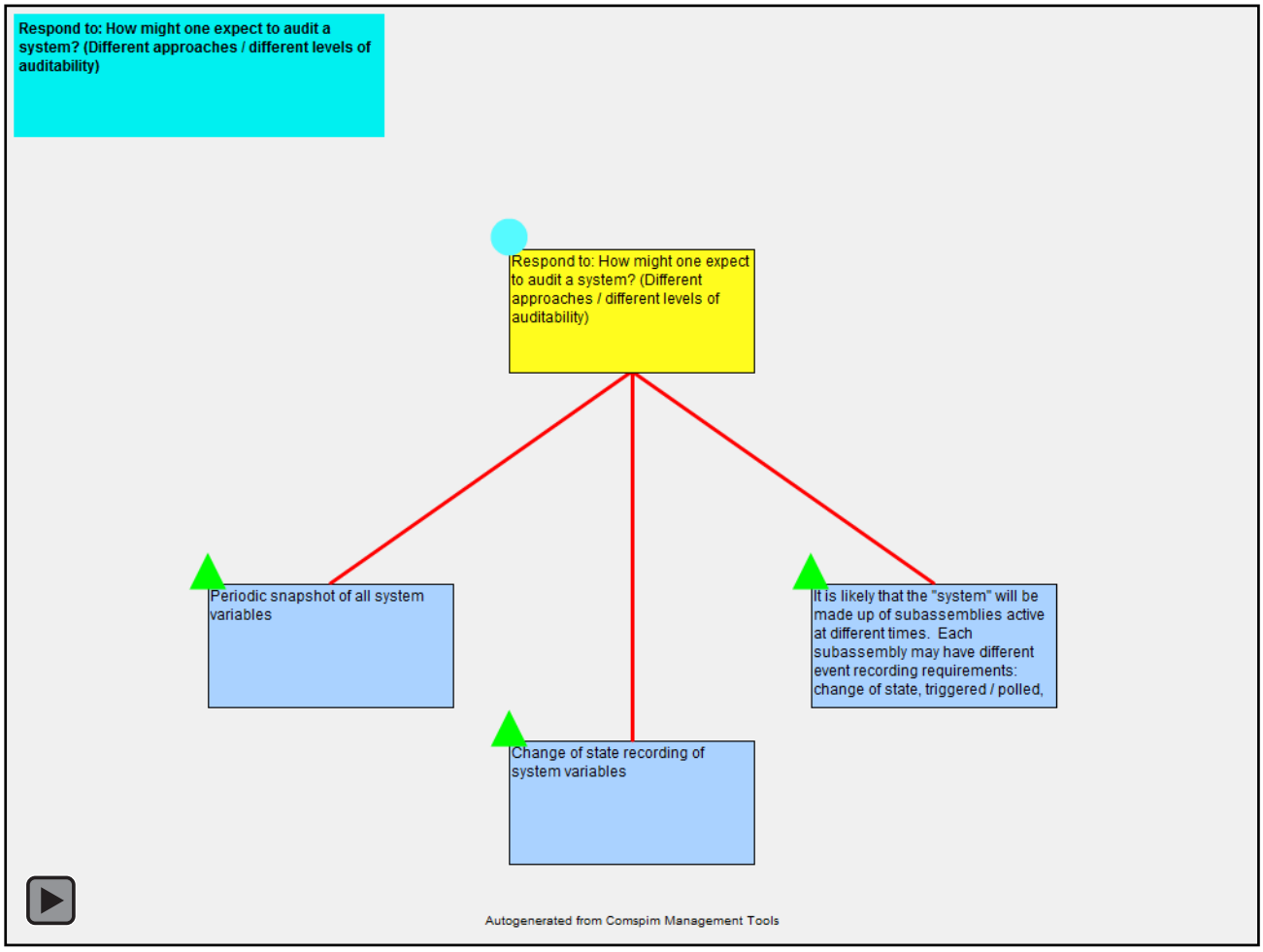

Figure 8

Full Auditability and Traceability Discussion Tree

A Requirement for Autonomous (Weapon) Systems?

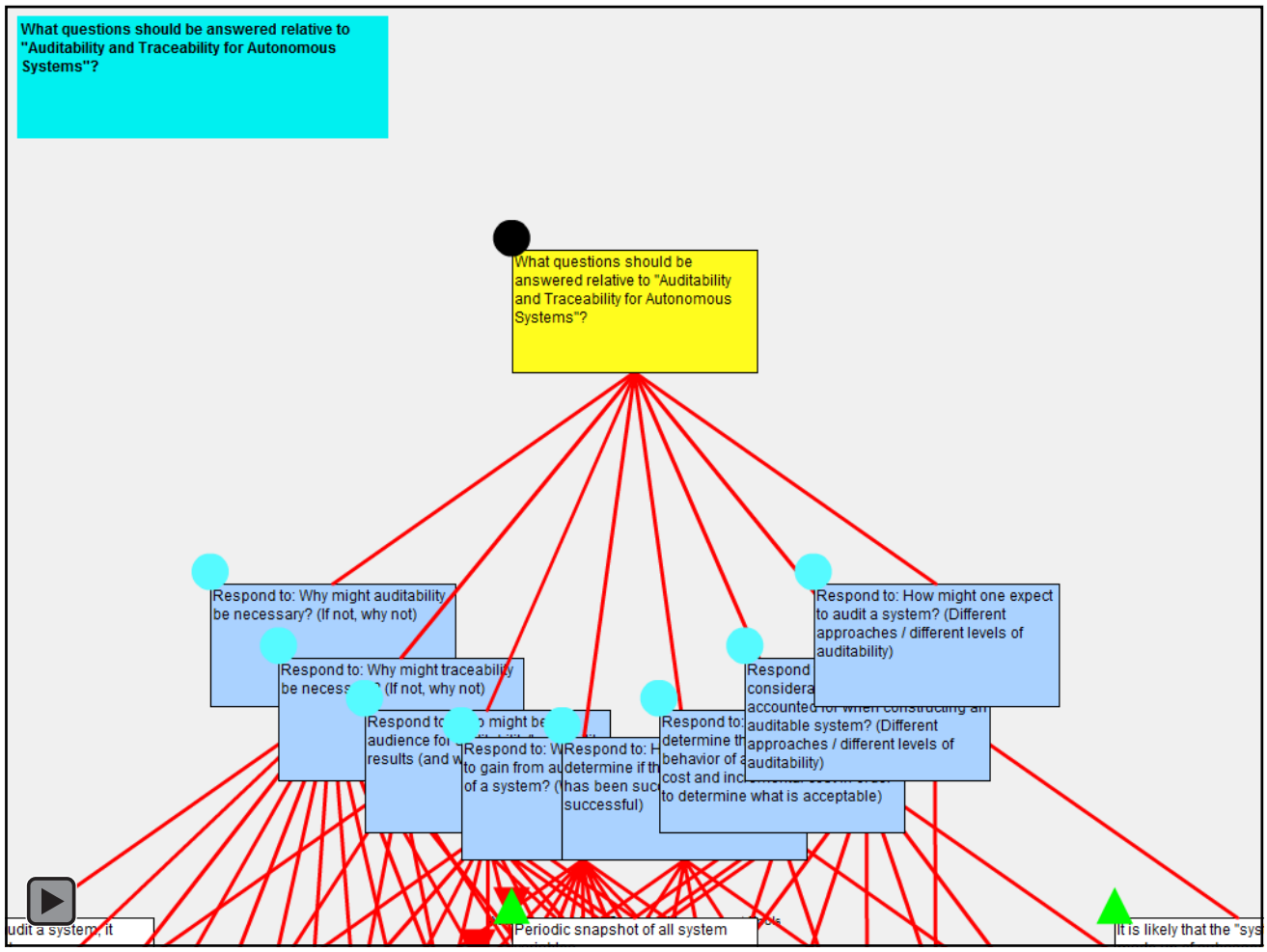

Figure 9

Appendix A

An interactive rendering:

The purpose of an interactive rendering is to enhance access to information so you can present more information in a smaller space and allow the user to interact with the information provided. This interactivity is supported with the keyboard arrow keys and the mouse.

When the interactive figure is first displayed, one box will be highlighted in yellow. This is the “selected box”. The text boxes can contain only a limited amount of text, so in the interactive figure, the full text of the “selected box” is displayed on the left side of the figure. In some cases this full text is hidden by other boxes. By moving the mouse over the blue area at the upper left corner of the figure, this full text view will be popped to the top.

Parents of the selected box will be highlighted in red. Children of the selected box will be highlighted in blue. This makes it easy to view the information hierarchy.

Navigation through the tree is accomplished with keyboard arrow keys.

Down Arrow will select the child of the current “selected box” (if one exists).

Up Arrow will select the parent (if any) of the current “selected box”.

Right Arrow will select the next box at the current level of the tree (if one exists).

Left Arrow will select the previous box at the current level of the tree (if one exists).

When the user navigates the information tree with the keyboard arrow keys, the new selected box will be positioned in the center of the display. In this way one can navigate a tree that has much more information than can be displayed in one screen.

You can change the selected box by clicking on a box with the mouse. In this case the box will be highlighted in yellow, but (unlike selection with the keyboard arrow keys) the box will not be automatically centered on the screen.

As you mouse-over a box (without clicking), it will pop to the top, and so will its parent box; thus making it easy to see information that would otherwise be hidden.

Appendix B

(Source Log from Compsim Management Tools)

What questions should be answered relative to "Auditability and Traceability for Autonomous Systems"?

1.1.

Respond to: Why might auditability be necessary? (If not, why not)

1.1.1.

Supporting Argument

If you cannot audit a system, it cannot be fixed.

1.1.2.

Supporting Argument

If you cannot audit its behavior, you cannot be sure it is doing what it is doing for the right reasons.

1.1.3.

Supporting Argument

If you cannot audit an autonomous system you cannot really understand it enough to extend it.

1.1.4.

Supporting Argument

If you cannot audit the behavior of a system you cannot identify weaknesses that might be exploited.

1.1.5.

Supporting Argument

You need to be able to audit a system so you can answer the question of "what was considered" and "what was not considered" in a decision or action.

1.1.6.

Supporting Argument

If you cannot audit the behavior of an autonomous system, you cannot demonstrate that it conforms to ROE and International Law.

1.1.7.

Supporting Argument

You need to be able to audit a system so you can see how information items influenced the decision or action, which would be necessary in developing a full understanding of a decision or action.

1.1.8.

Supporting Argument

If you cannot audit the behavior of an autonomous system, you cannot debate its existing policy.

This requires that one is able to understand the format of the audit.

1.1.9.

Supporting Argument

Auditability may be necessary so the policy development process can be used in training future policy makers.

1.1.10.

Supporting Argument

From my standpoint as a weapons and combat systems developer, I need to have systems that can tell me why they did what they did.

If something then happens that didn't fit my idea of what should have happened, I can go back into the system and request the audit trail that led to a particular action to see where the system went astray. This gives me a place to start looking for what needs to be done differently so that we don't have the same thing happen again.

1.1.11.

Objecting Argument

If an Autonomous System is captured by opposition, there may be a need to erase all behaviors to keep that policy from opposition forces.

1.1.12.

Objecting Argument

There may be cases where an organization wants to perform clandestine operations and they don't want to have their actions reviewed.

1.1.13.

Objecting Argument

In extremely small Autonomous Systems, there may not be sufficient hardware resources to retain information to support auditability.

1.1.14.

Objecting Argument

There will be no need to audit every decision and action that an autonomous system makes in a production system.

While during initial development and testing, every decision or action will need to be reviewed, once fielded, there may be interest in auditing only critical decisions and actions.

1.2.

Respond to: Why might traceability be necessary? (If not, why not)

1.2.1.

Supporting Argument

When something goes wrong, it is important to trace the problem to the appropriate person.

It is important to note that sometimes things go wrong for unexpected reasons. It is also important to note that probably everything will fail in some way at some time under some circumstances.

1.2.2.

Supporting Argument

When systems perform "per policy" then it is important to understand who is responsible for the policy.

1.2.3.

Supporting Argument

It is likely that some parameters used by an autonomous system may be adjusted by humans in the field for specific scenarios. Those adjustments will affect the behavior of the system. Those adjustments need to be traced to that human.

1.2.4.

Objecting Argument

Traceability (to a responsible person) will not be necessary for some decisions or actions executed by an autonomous system.

For example, politically trivial actions: The autonomous system decides to go right around an obstacle rather than left. In this case and engineer may want to audit the system to understand why, but traceability to the person that created the policy that caused the autonomous system to choose the path may not be important.

1.2.4.1.

Is it true for military systems, or is traceability to every line of code important?

1.2.5.

Objecting Argument

There may be a desire to eliminate traceability to conceal the senior personnel responsible for a policy.

1.2.6.

Objecting Argument

There may be a desire to conceal responsibility by a person that makes a configuration adjustment that changes the behavior of the autonomous system so that person can avoid responsibility.

1.3.

Respond to: Who might be the audience for auditability/traceability results (and why)?

1.3.1.

Supporting Argument

Auditability: The policy maker needs the ability to audit the behavior of the AxS to insure it performs as desired.

1.3.2.

Supporting Argument

Auditability: The Test Engineer needs to audit the behavior of the system to insure it performs as expected.

1.3.3.

Supporting Argument

Auditability: The field commander needs to perform after mission reviews to review safety critical decisions made by the AxS.

1.3.4.

Supporting Argument

Auditability: The field service technician will need to audit the behavior of the system to trace failed system components.

1.3.5.

Supporting Argument

Auditability: The manufacturer will need to audit the behavior of the system to evaluate design issues.

1.3.6.

Supporting Argument

Auditability: Lawyers will need to audit the behavior of the AxS should legal issues with behavior of the system be questioned.

1.3.7.

Supporting Argument

Traceability: The lawyers will need traceability to assign responsibility if inappropriate "legal / ROE" decisions are made by the AxS.

1.3.8.

Supporting Argument

Traceability: Politicians / Government Leaders will need traceability to assign responsibility.

Black swan events: when an unexpected event happens and millions of dollars are spent tracing responsibility to the source.

1.3.9.

Supporting Argument

Individuals responsible for "system safety" will need the ability to audit the behavior of autonomous systems.

1.4.

Respond to: What might one want to gain from auditing the behavior of a system? (What information)

1.4.1.

Supporting Argument

In complex systems it will be important to build trust in the system and the embedded policies, so it will be important to understand why they perform actions, and perhaps more important, why they do not do expected actions.

It is likely that policies for Autonomous Systems evolve over time and get better and better as new challenges are encountered. These challenges will identify weaknesses in the policies, necessitating new sensors, data sources, new policy influences.

1.4.2.

Supporting Argument

Understand how each of the information items was valued at the time of a critical decision.

1.4.3.

Supporting Argument

Identify missing pieces of information that might have been "implied" when making decisions.

1.4.4.

Supporting Argument

Identify the timeliness of information used in making critical decisions and how this information value had degraded over time (if that is provided).

1.4.5.

Supporting Argument

A specific list of information items that were included in the making a decision or controlling an action.

May identify obsolete or incomplete policies when audited.

1.4.6.

Supporting Argument

Identification of conflicting information with a view into how conflicting information was handled.

1.4.7.

Supporting Argument

Indication of trust assigned to information used in decisions or actions.

1.4.8.

Supporting Argument

Sources of information used in decisions and/or actions.

1.4.9.

Supporting Argument

Authors of the Policies or Policy Components that might want to be held accountable.

1.4.10.

Supporting Argument

Configuration parameters that contribute to the policy that control behavior.

1.4.11.

Supporting Argument

Time stamp and records associated with adjusting configuration parameters.

It is likely that some policy parameters are configurable. Different humans throughout the weapon system hierarchy will likely have different configuration privileges. It will likely be important to know who configures what and when.

1.5.

Respond to: How might one determine if the auditing process has been successful? (or can be successful)

1.5.1.

Supporting Argument

After mission reviews can be conducted to insure decisions and actions by the Autonomous System were correct.

1.5.2.

Supporting Argument

System Safety reviewers are able to satisfy themselves that the autonomous system performed correctly and within defined boundaries.

1.5.3.

Supporting Argument

The system designers are able to validate their design during the development process.

1.5.4.

Supporting Argument

The policy makers can validate the performance of their policies.

1.5.5.

Supporting Argument

Anyone can easily trace decisions and actions to the source / cause / justification.

1.5.6.

Supporting Argument

The Lawyers can understand the value system used by the autonomous system and understand the policy implemented by the autonomous system, with sufficient detail to argue the merits of the decisions or actions.

1.6.

Respond to: How might one determine the cost of auditing the behavior of a system? (Recurring cost and incremental cost in order to determine what is acceptable)

1.6.1.

Supporting Argument

Recurring Cost: Hardware / memory costs for retaining information sufficient to audit the behavior of the autonomous system.

1.6.2.

Supporting Argument

Recurring cost: Hardware and software costs associated with extracting recorded information used to audit the behavior of the autonomous system.

1.6.3.

Supporting Argument

Incremental cost: Human time and effort required to review and understand recorded information that drove the behavior in decisions and actions of interest.

1.6.4.

Supporting Argument

Incremental cost: Training of humans that will audit the autonomous systems in the methodology used to package the behavior being audited.

1.7.

Respond to: What system considerations need to be accounted for when constructing an auditable system? (Different approaches / different levels of auditability)

1.7.1.

Supporting Argument

Performance impact to record information necessary for an audit.

1.7.2.

Supporting Argument

Selection of which decisions and actions need to be audited in a development environment.

1.7.3.

Supporting Argument

Selection of which decisions and actions need to be audited in a production environment.

1.7.4.

Supporting Argument

Selection of triggers that will record information of decisions and actions that needs to be audited.

1.7.5.

Supporting Argument

Specification and agreement on limitations of what will be recorded.

It will not be possible to record "everything" "all the time" as an autonomous system will have limited memory retention space.

1.7.6.

Supporting Argument

Specification of mandatory event recording and running recording for an autonomous system.

It is likely that auditing some decisions and actions will be mandatory (such as specific engagement decisions where weapons are used). Resources must be included for these events.

1.7.7.

Supporting Argument

Some systems may require that reasoning / justification for decisions and actions be concealed from review under certain circumstances. So some means of flushing memory may be required.

Example: Erasing history from one mission to the next. Or erasing history when a policy upgrade is performed.

1.8.

Respond to: How might one expect to audit a system? (Different approaches / different levels of auditability)

1.8.1.

Supporting Argument

Periodic snapshot of all system variables

1.8.2.

Supporting Argument

Change of state recording of system variables

1.8.3.

Supporting Argument

It is likely that the "system" will be made up of subassemblies active at different times. Each subassembly may have different event recording requirements: change of state, triggered / polled, time scheduled, priority event triggered.