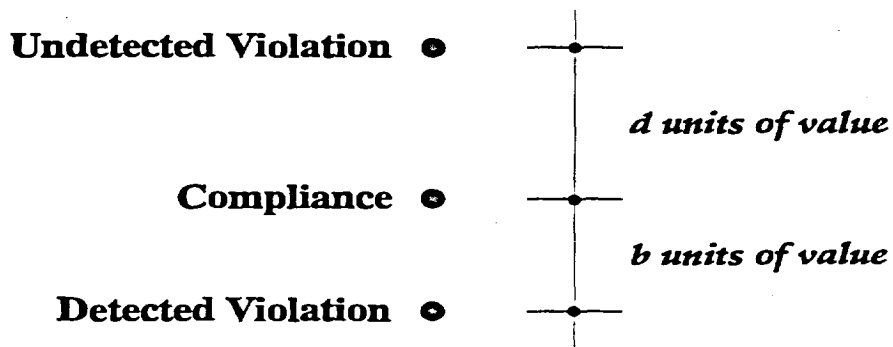


values will always be measured relative to its value for compliance. Thus, the indication "State receives  $d$ " beside the Undetected Violation outcome in Figure 1a means that the state expects that for an undetected violation it would gain  $D$  units of value more than what it would gain for Compliance. Similarly, "State receives  $-b$ " at the Detected Violation outcome means that the state anticipates that it would lose  $b$  units of value if a violation were detected, relative to what its position would be if it chose Compliance.

In summary, the state's values are as described in the following diagram:



In other words, compliance represents the status quo level, against which all gains and losses are measured. (Technically, value differences are measured in von Neumann-Morgenstern utilities — see [4] for details.)

Clearly, the notation assumes a positive gain for an undetected violation. Under the NPT, this is almost surely not the case for most states most of the time. But note that if an undetected violation is worth less than compliance, then nothing is ever gained by violating, so there is no compliance problem, and safeguards are actually unnecessary. The assumption made here — that an undetected violation improves the state's position — means that the model of Figure 1a, and all of the models below, address situations in which there is a potential compliance problem. Likewise, the notation implies that a detected violation is worth less than compliance. Without this assumption, there is never any reason for the state not to violate. Thus the relative values implied by the notation and described schematically above ensure that the model does not address situations in which a safeguards program is either unnecessary or infeasible.

How does a state make its choice when faced with a decision problem such as the one shown in Figure 1a? The state must evaluate the consequences of each of its alternatives, Violate or Comply. To do this, it must include in its assessment not only the values of the possible consequences of choosing Violate, namely Undetected and Detected, but also it must include the