Improving Performance in Detection Systems: Exploring a Parallel-judgment Structure

Ignacio J Martinez-Moyano, Ph.D.
Argonne National Laboratory

Abstract:

In this paper, we describe a model of two judges in parallel. We expand work related to the identification of threats by analyzing the effectiveness of including a second judge in the process and identifying its effects on performance and error rates. In addition, we explore the implications for learning under uncertainty.

Introduction

Improving performance in detection systems is one important challenge in dealing with terrorism and insurgencies. Terrorist activities are based on covert operations and make use of deception to hide intent and possible detection before action occurs. The intelligence community and law enforcement personnel struggle to improve detection systems to counter activities of terrorists and insurgents across the world. In the decision sciences and learning disciplines, most notably in the cases of cyber threats and attacks, selection-detection processes have been studied as part of an effort to increase the accumulated knowledge of how to detect those elements that belong to a distribution of interest (signal) when such elements are embedded in a larger population of elements that are not of interest (noise).

The challenge of identifying the members of a distribution of interest (signal) in a certain population belongs to a broad class of problems known as selection-detection problems. Frequently, this class of problems is addressed with signal-detection theory (Egan, 1975; Green & Swets, 1966; Swets, 1992). In this type of problem, a mechanism needs to be used to make a judgment concerning the likelihood that the element belongs to either the signal or noise distribution to determine appropriate action (e.g., keep or discard, let through or stop, arrest or not, accept or reject). The said mechanism can be automated or human-based. The performance of the judge (automatic or human) determines the adequacy of the detection-selection process. A perfect detection-selection process should be capable of identifying the members of both distributions (signal and noise) without errors. Figure 1 shows how the two distributions typically overlap each other, making the matter of selection-detection anything but trivial.

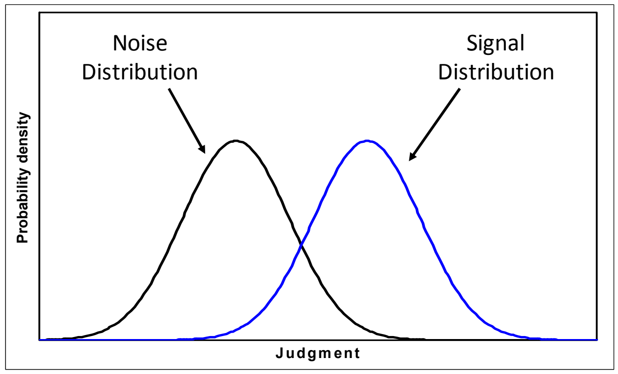


Figure 1—Selection-Detection Problem Characteristics

A selection-detection problem is trivial when the targeted population (signal) is clearly distinguishable from the non-targeted one (noise). Figure 2-A presents a graphical representation of this case: the no-uncertainty case. In this case, the two distributions do not overlap, allowing for a level of judgment (along certain dimensions or a composite of dimensions) that clearly determines where one distribution ends and the other starts. The no-uncertainty case exists in the following example: the task, in a population of 100 items composed of squares and circles whose shapes are fully visible, is to separate (detect and select) the squares from the circles. The distribution of circles and the distribution of squares are clearly distinguishable and, therefore, a judge (human, mechanical, or electronic) could easily separate the two types of items without any errors. In the context of the identification of terrorists, this challenge would be equivalent to identifying a characteristic of individuals such that terrorists would be unmistakably and clearly identifiable from other individuals. (If such a characteristic or bundle of characteristics exists, it is not yet known, making the identification of terrorists an important challenge). In reality, most detection-selection problems are complex and difficult to address. Judges have to rely on multiple, fallible information cues about the distributions to attempt to distinguish members of the signal from the noise (Hammond, 1996, 2000). The detection-selection problem becomes more difficult as the area of overlap of the distributions increases. Figure 2-B shows a graphical representation of a small level of uncertainty in this type of problem. As the figure shows, the two distributions overlap, creating the possibility of error. Figure 2-C shows a higher level of uncertainty by presenting a larger area of overlap of the distributions. As shown in these figures, the larger the area of overlap, the more difficult it is for a judge to distinguish which elements belong to which distribution. Thus, a judge making a judgment under conditions of uncertainty in a detection-selection task will unavoidably face the possibility that two types of error will occur: false positives and false negatives.

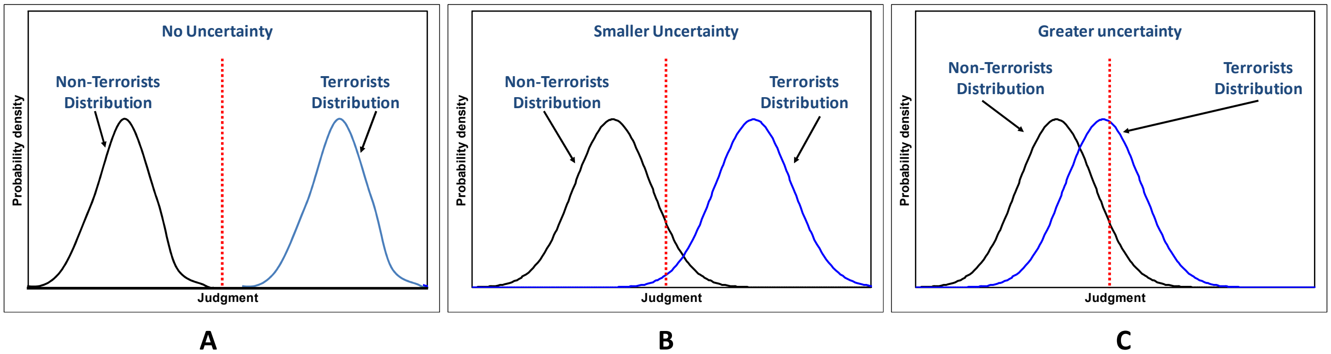


Figure 2—Uncertainty Configurations

A judge makes a false-positive error when he, she, or it mistakenly identifies a member of the noise distribution as belonging to the signal distribution. This error is also identified as a false alarm. Depending on the context, this type of error might not be as costly as a false-negative error. False-negative errors happen when a judge (or a selection-detection system) mistakenly identifies a member of the signal distribution as belonging to the noise distribution. In the context of terrorism, this type of error equates to allowing a terrorist through a security checkpoint (at an airport or border crossing point) instead of arresting the individual, thus allowing him or her to carry out his/her mission. False-negative errors tend to be more costly than false-positive errors. In the context of terrorism, a false-negative error can have devastating consequences, generating a large amount of economic damage and casualties. Therefore, in this context, judges normally tend to err on the side of incurring more-than-needed false-positive errors to prevent false-negative errors from occurring. Adding to the complexity of the problem, some terrorists are deliberately trying to sneak through defenses to learn about the defense systems and to be able to report back to their groups. In these cases, terrorists do not exploit the penetration of the defenses immediately, making it more difficult for judges to identify when errors happen. In this sense, and not only in the context of terrorism, many detection-selection processes are forced to operate under conditions of limited and conditional feedback (Ghaffarzadegan, 2008).

In many countries, and in particular in the United States, selection-detection processes with an important human-based judgment component are widely used in the prevention of terrorism. In particular, and among many others, the Transportation Security Administration and the Bureau of U.S. Citizenship and Immigration Services of the Department of Homeland Security (DHS) deploy tens of thousands of individuals (judges) every day to selection-detection posts at airports, maritime ports, roads, and border-crossing points who are charged with the task of identifying and stopping malicious individuals from harming the people and institutions of the United States. In general, these judges have the ability to make a call individually but also rely on the assessment of their peers to conduct their very important work. Policies and procedures are in place to increase the probability that they will successfully carry out their detection and selection tasks.

We have elsewhere explored the fundamental components of iterative detection-selection processes in a single-judge structure and have identified these components to be judgment, learning, and decision-making processes (Martinez-Moyano, Conrad, & Andersen, 2007; Martinez-Moyano, Rich, Conrad, Stewart, & Andersen, 2006). In this paper, we use formal modeling to learn about the effects that parallel judgment has on the performance of detection-selection processes. In order to explore this idea, we are expanding on previous work and building a dual-judge structure in which the judgment process occurs in parallel, thereby introducing a multi-view of the detection problem. Expanding previous research (Martinez-Moyano, Rich, Conrad, Andersen, & Stewart, 2008), in this model we represent a situation in which two judges are exposed to the same set of information cues about a certain phenomena and they (together) become a higher-level judge that decides about the likelihood that a case belongs to either the signal or noise distributions.

In the next section, we explain the main parts of the structure of the model that we use to explore the dual-judge system. Later, we present the simulation scenarios used in this investigation and, in the last part of the paper; we discuss the findings, present conclusions, and advance our thinking on future avenues for research.

Model Structure

Judgment Process

The dual-judge model is organized around three main processes: judgment, decision making, and learning. In this model, we integrate constructs from social judgment theory (Brunswik, 1943, 1956; Hammond, 1996; Hammond & Stewart, 2001), signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 2005), and outcome-based learning (Erev, 1998; Erev, Gopher, Itkin, & Greenshpan, 1995) in a system dynamics framework (Forrester, 1961; Richardson & Pugh, 1981; Sterman, 2000) to capture the process in which two judges make assessments about a series of cases with respect to their likelihood of representing a threat, thus triggering defensive action.

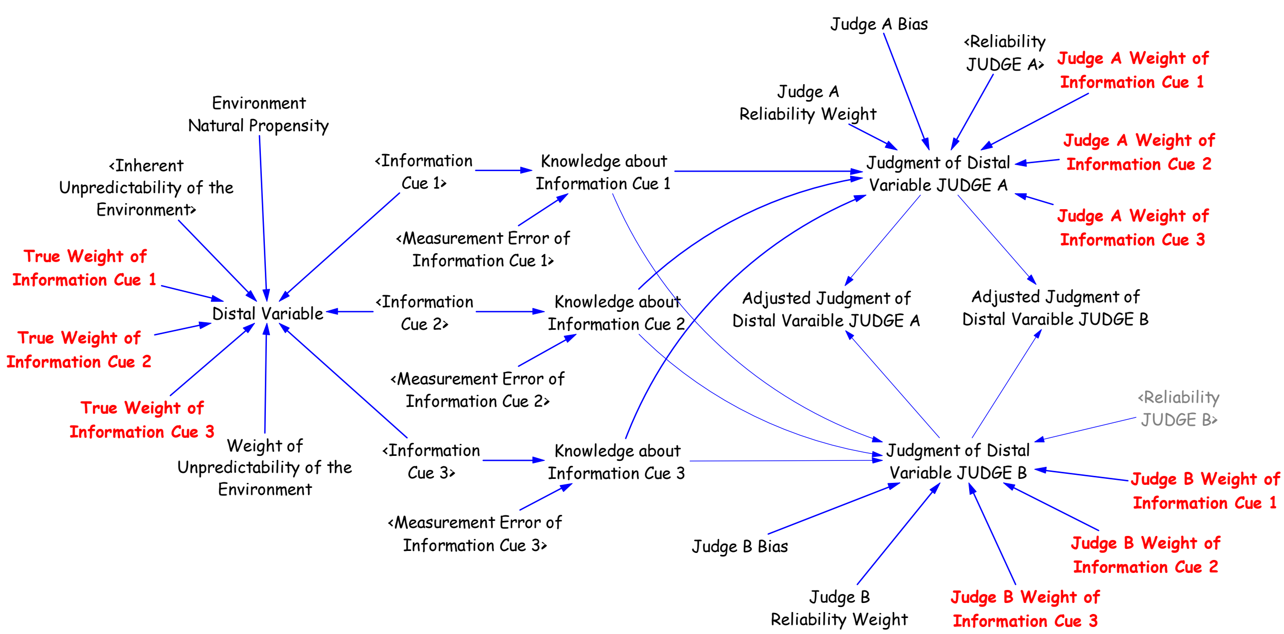


Figure 3—Judgment Process

In Figure 3 (above), we show how we structure the judgment process in the model. In the model, two judges (Judge A and Judge B) make judgments about the nature of a distal variable. Both judges use the same information vector (information cues about the distal variable) and have the same judgment policies (bias, reliability, and relative weights for information cues). In this sense, the two judges are symmetrical (later, we will relax this condition). The symmetry of the judges is used to represent an extreme case — desirable to many in the security industry — in which all judges present in a security checkpoint share the exact same policies (in practice, however, no two judges have identical judgment policies). We advance the identical-judges scenario as desirable given our experience in identifying high levels of training and on-the-job mentoring as pervasive and highly regarded in many agencies as a good way to get officers and inspectors to improve their judgment skills and to ensure minimum standard levels of practice.

As explained earlier, the judges share the exact same information vector about the environment. In addition, to relax the full-symmetrical assumption, in the model we introduce a two-stage judgment mechanism that operates as follows: first, we allow the judges to integrate information about the environment and create a judgment of what they think the distal variable is. Once the "first" judgment is produced, the judges produce an "adjusted" judgment in which two information cues are used: their own judgment of the distal variable (the "first" judgment) and information about the other judge's judgment of the distal variable. The judgment process of each judge is characterized according to social judgment theory as a function of the information cues about the phenomenon under study and a number of parameters that capture the individual characteristics of the judge. (In Table A and Table B of Appendix 1 of the paper, we present these parameters and their base values). In general, the judgment process is characterized as a weighted average of the information vector available to the judge and complemented with bias and reliability parameters. The judgment equation is of the form:

$$Y = \hat{Y} + e$$

Where Y is the judgment of the distal variable, $\hat{Y}$ is an estimate of Y, and e is an indicator of degree of reliability of the judge (and, in the model of the environment, represents the inherent unpredictability of the environment).

$\hat{Y}$ captures the judgment process by combining information about the distal variable and judgment bias. $\hat{Y}$ is of the form:

$$\hat{Y} = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + k$$

Where $\hat{Y}$ is an estimate of Y, b_n is the weight of information cue *n* on the judgment, X_n is the information cue *n*, and k is a bias term.

Using a two-stage judgment process, we model the possibility of the judges' mutual influence in the detection-selection task. This element is an important part of the cognitive process that multiple individuals (e.g., agents, monitors, judges, screeners) face on an everyday basis. When a two-stage judgment process is institutionalized, it provides the basis for teamwork synergy to emerge. In many contexts, including law enforcement and the military, a multi-judgment environment is fostered to ensure that adequate calls are always made. For example, in difficult cases in medical radiology, it is common to have a tiered system in which multiple physicians independently examine the same body of evidence (image) and then make a judgment as to what it is that the image shows. Once these multiple opinions are generated, the physicians share their views, creating the opportunity to adjust their assessments by incorporating additional information coming from the expert judgment of their peers. In this way, junior radiologists learn about the process and find ways to fine-tune their abilities by seeing what senior radiologists do. In previous work, we have explored the simpler one-stage judgment process and found it to be a fair representation of what judges do in detection-selection tasks in experimental settings (see Martinez-Moyano et al., 2007) when working alone. However, very important detection-selection processes are carried out by judges working in groups (formally or informally) on a daily basis, making the need for insights about this process salient.

Additionally, in this section of the model, we characterize the environment that is being presented to the judges. We model the judgment process and the environment in a symmetrical way. Both models share the same structure but are parametrically different. In this way, in the model we have information about the environment that we use to compare to what the judges are able to identify and integrate. In our model, we have information about the true state of the environment and about what the judges perceived of the environment. With this information we can measure judgment accuracy and overall performance of the detection-selection process. In particular, we vary the inherent predictability of the environment to assess the benefits of having a dual-judge structure and other judgment policies. The “Weight of Unpredictability of the Environment” captures the inherent predictability of the environment in the model. A zero (0) weight represents a perfectly predictable environment with no stochasticity (and formulated linearly). A weight of 1 represents an environment that is as unstable as it can possibly be, given a certain distribution of stochastic noise in its predictability. We model an environment with a low base rate of incidence of signal (13.6% parametrized in the model via a criterion threshold of 61 in a normal distribution with a minimum of 0, a maximum of 100, a mean of 50, and a standard deviation of 10).

The judges, therefore, integrate information about the environment and about other judges’ judgments to create a judgment that is then compared to a decision threshold that determines whether defense action is granted or not. When their judgment of likelihood is higher than their decision threshold, action is triggered. Thus, action is the result of the coupling of judgment and decision making.

Decision-making Process

After judgments are generated, these need to be used in a decision-making process to determine whether their level is sufficient to grant action. This process is captured in the model by comparing the level of the adjusted judgment with that of a decision threshold that determines the cutoff of what is acceptable and what is not (see Figure 4).

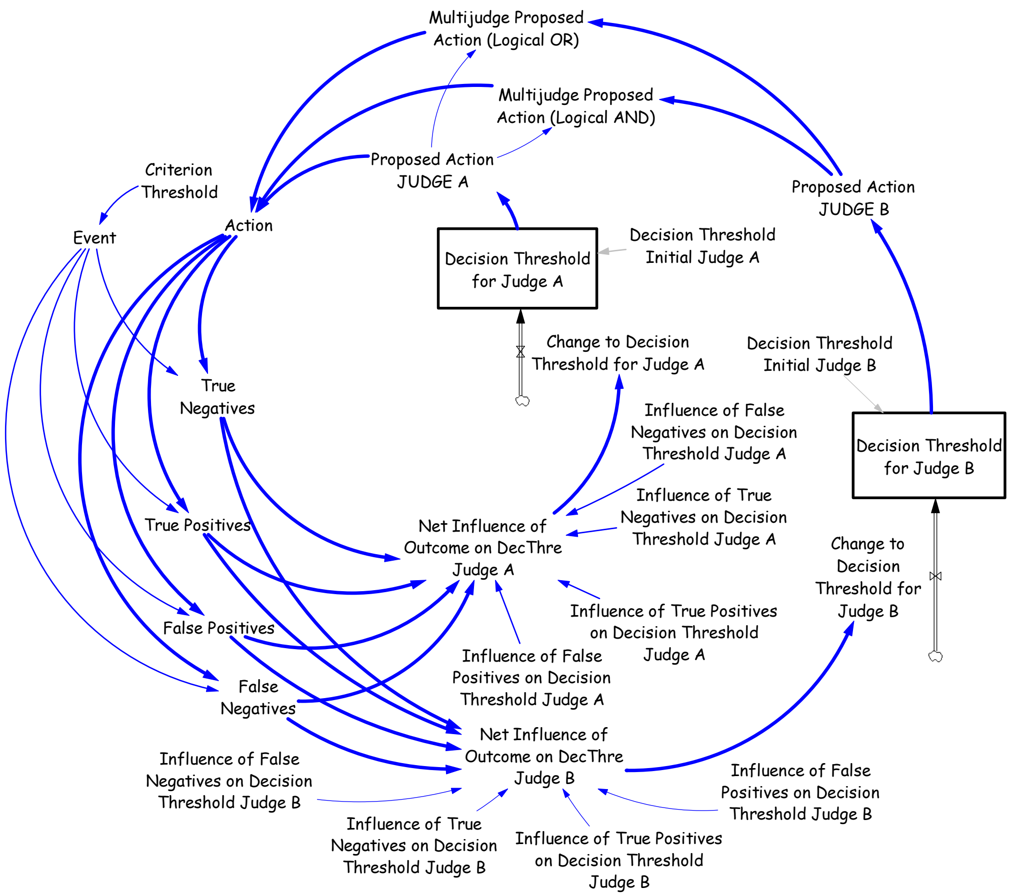


Figure 4—Decision and Learning Processes

Each judge has his or her own decision threshold that he/she uses for action-related determinations. In this way, the decision threshold is as instrumental in the determination of action as the judgment is. The coupling of these two elements is what allows judges to infer whether action is needed or not. After decisions are made and outcomes are experienced, the potential to update the decision thresholds exists. For example, in the context of the identification of terrorists, let us suppose that an agent evaluating cases assesses a case with an index of 45 (in a 100-point scale). Should this raise red flags and trigger action? The answer will depend on the decision threshold that determines what is "normal" and what is not. If in this case, the agent has a decision threshold of 60 points, then the case is considered "normal" and no action is granted. However, if the agent operates with a decision threshold of 35 points, then the case is above the cutoff point and should be acted upon. In this sense, the decision-making process is very straightforward and clean. However, several potential problems exist that contaminate the process, making certainty problematic (or inexistent). First, the original 45-point assessment is based on fallible information cues and is subject to imperfect judgment and biases. Also, the decision threshold level (60 or 35) is also subject to biases and uncertain indicators based on outcome identification. Therefore, if the judge with the 60-point threshold does not engage in defensive action, he might incur error in letting a terrorist walk. Alternatively, when the judge with the 35-point threshold intervenes to stop the suspect, he might be detaining an innocent individual. Further, the judge might make the correct call for the wrong reasons creating a potentially disastrous learning environment. For example, let us suppose that the original 45-point assessment was wrong and that the true level was 65 points. Also, let us suppose that the judge mistakenly used the 60-point threshold when in fact he should have used a 75-point threshold that matches the base rate of occurrence of the phenomenon. In this case the judge has incorrectly assessed the case and has incorrectly updated her decision threshold while at the same time, by not engaging in a defensive action, the judge is choosing a correct course of action (as the real assessment is lower than the real threshold). As the outcome of this decision is successful, the judge will incorrectly reinforce his beliefs about how to assess cases and how to update his decision threshold. Accidental success becomes the origin of superstitious (and quite mistaken) learning.

Learning Process

Independent of what type of error is more costly (false positives vs. false negatives); minimizing error happens through a fundamental learning mechanism that links outcomes of the decision process with updated levels of the decision threshold (Erev, 1998). As outcomes are experienced and errors are recognized, judges modify the level of the decision threshold to find the level that creates higher levels of performance. This is not the only learning mechanism that judges use to improve performance. Other mechanisms, which are not present in the model, include improving: the judgment process (including reliability, biases, function forms, etc.), selection technology (i.e., pushing the signal and noise distributions further apart), and quality of the information received about the environment, etc.

Updating the decision threshold is complicated for judges, because feedback from their decisions is sometimes difficult to obtain and in many cases is not close in time or space to the original decision (Weaver & Richardson, 2002). In the model, we use an outcome-feedback learning process to update the judge's decision threshold. In this sense, judges learn by doing in repeated iterations. Each judge has his or her own value matrix that he or she uses to modify the decision threshold. A value matrix captures the value that each type of outcome has for the judge. Correct and incorrect decisions yield results that the judge identifies and uses to adjust the decision threshold. In general, the absence of errors reinforces the use of the current decision threshold, while recognition of errors causes the judge to change the level. Depending on the context, the different types of error might have very different values and effects on changes to the decision threshold.

Additionally, in the model we use a number of flags that allow us change configurations of the judges and of the judgment and decision-making processes to identify conditions that have the potential to improve performance of detection-selection processes. The most important flags used in the model are explained in Table C of Appendix 1. Additional details of the learning theory implemented in this model can be found in (Martinez-Moyano et al., 2006).

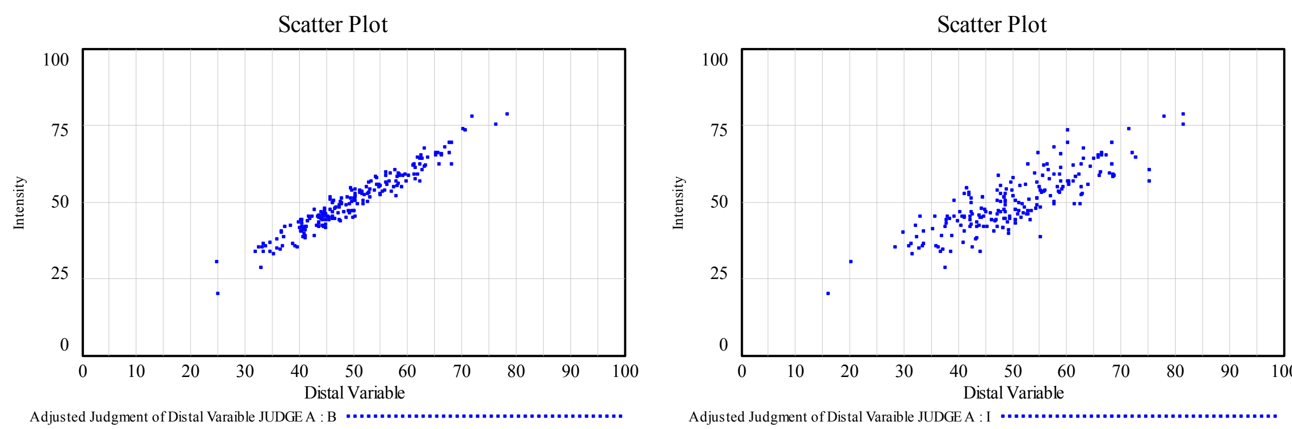
Simulation Scenarios

In order to understand the impact of changes to the characterization of judges, the judgment process, and the environment on the performance of the detection-selection processes we modeled, we designed 34 different scenarios to be used. For all simulated scenarios, the base rate of incidents remains constant (13.6%). In Table 1, we show these scenarios, the variables that we use to characterize them, and the values used in the different simulations. We use the model to learn more about the impacts that the different symmetry levels of the judges, different conditions of the environment, different rates of learning, and different conditions of coordination among judges have on performance metrics of the detection-selection process.

The predictability of the environment is characterized at two levels: perfect and unstable. A perfect environment is one that is fully predictable (i.e., no stochasticity is built in), and an unstable environment is 50% predictable (i.e., it is modeled by using a stochastic distribution around the perfect environment). We anticipate that modifications to the inherent predictability of the environment will have an impact on process performance as articulated in proposition 1.

Proposition 1: Under conditions of perfect predictability of the environment, lower error rates and better performance will be achieved.

In Figure 5, we show scatter plots of the judgments of Judge A relative to the actual measure of the distal variable for runs B (perfect environment) and I (unstable environment). The judgments of a perfectly reliable judge would correlate better with the environment in both cases. Note that the spread of the judgments is higher in Figure 5-B than in Figure 5-A.



A B

Figure 5—Dispersion of Judgemnt

The frequency of learning is varied at three levels: normal, high, and low. Frequency of learning is contolled with a sampling mechanism that creates the condition for different levels of exposure to the results of the judgment process. In high-frequencey cases, judges receive four times as much feedback as in the normal case. In the low-frequency case, they rceive one-fourth of the feedback they would normally receive. The differing of frequency levels captures the possibility of having different frequencies of feedback provided to judges and estimating what impact(s) that would have on performance and error rates. We anticipate that frequency of learning will influence error rates and performance levels as articulated in proposition 2.

Proposition 2: Under conditions of higher frequency of feedback, lower error rates and better performance will be achieved.

Judgment symmetry is varied by means of judge characterization parameters related to judgment policies, function forms, bias, initial decision thresholds, and judgment reliability. We create two basic conditions: symmetrical judges and asymmetrical ones. Symmetrical judges are identical in every way, including in their level of judgment reliability. The judges in the model are not perfectly reliable (although they do have the same level of unreliability in judgment). Asymmetrical judges differ in judgment reliability (i.e., the reliability of Judge A is higher than the reliability of Judge B) and in the initial levels of their decision thresholds (i.e., Judge A starts with a higher decision threshold [56] than Judge B [25]). The parametrization of asymmetrical judges creates scenarios in which two distinct judges interact. Judge A becomes the more reliable judge of the two, and Judge B becomes the more prudent at the start of the simulation by having a lower initial decision threshold. Judgment reliability is desirable in judgment processess, and judgment prudency (i.e., having low decision thresholds) ensures the avoidance of critical errors. In real detection-selection systems like those set up across the United States (and in many other countries) to prevent terrorists from gaining access to critical infrastructure (such as through air travel), thousands of human judges engage their abilities every day to protect the public. In these systems, it is highly likely to find several different judges interacting, some who are more reliable than others and some who are more prudent (i.e., less tolerant to uncertainty) than others. Our characterization of the asymmetrical judges in the model is designed to capture these differences in a combined manner. It is clear to us that a different combination of judgment characteristics can be identified and codified for judges. Also, we anticipate that changing these characteristics may change the results presented here. However, we are confident that using the parameters we chose will allow us to identify certain conditions under which asymmetry in judgment may benefit/hinder performance in detection-selection systems.

Additionally, in order to be able to capture the value of having more than one judge in the decision process, we introduce a coordination parameter that links results of the pairing of judgment with decision tresholds and action triggering. As in the dual-judge model, we have two potentially different opinions related to launching a defense action, and we created two conditions, or policies, on how to determine whether action is granted. The first one is the no-agreement-is-needed policy. In this policy, when either one of the judges determines that defensive action is needed, an action is launched. Alternatively, in the agreement-is-needed policy, defensive action is launched only when both judges agree that it is needed. This second policy imposes a more stringent control on action triggering. We anticipate that symmetry in judgment and coordination in decision making will have an effect on performance as articulated in propositions 3 and 4.

> *Proposition 3: Decision-making coordination will yield lower error rates and better performance independently of judgment symmetry.*

> *Proposition 4: Under conditions of asymmetrical judgment, higher variability of error rates and performance will be observed.*

Lastly, in order to test the effect of the two-stage judgment process, we designed scenarios in which symmetrical and asymmetrical judges use information about the other judge in their judgment process to create adjusted judgments before determining whether action is needed (by comparing the adjusted judgment with their decision threshold). These scenarios (runs AA through HH) are explored in both environmental conditions (perfectly predictable and unstable) but only at the normal frequency of learning. Coordination of decision making is also varied in these conditions. We anticipate that the use of a two-stage judgment process will have an impact on performance as articulated in proposition 5.

> *Proposition 5: Under conditions of a two-stage judgment process, lower error rates and better performance will be achieved.*

Table 1—Scenario Characterization

#	ID	Run Name			Variables										
			Environment	Learning Frequency	Weight of Unpredictability of the Environment	Judges Involved Flag (Dmnl)	Agreement for action needed Policy Flag (Dmnl)	Judge A Reliability Weight (Dmnl)	Judge B Reliability Weight (Dmnl)	Frequency of Sampling (Period)	Same Size Change for Judges Flag (Dmnl)	Decision Threshold Initial Judge A (Intensity)	Decision Threshold Initial Judge B (Intensity)	Judge A Weight of Judgment of Judge B	Judge B Weight of Judgment of Judge A
1	A	Base	Perfect Environment		0	1	0	0.25	0.25	1	1	56	56	0	0
2	B	Symmetrical Judges (Agreement not Needed)		Normal	0	2	0	0.25	0.25	1	1	56	56	0	0
3	C	Symmetrical Judges (Agreement Needed)			0	2	1	0.25	0.25	1	1	56	56	0	0
4	D	Symmetrical Judges (Agreement not Needed)		High	0	2	0	0.25	0.25	0.25	1	56	56	0	0
5	E	Symmetrical Judges (Agreement Needed)			0	2	1	0.25	0.25	0.25	1	56	56	0	0
6	F	Symmetrical Judges (Agreement not Needed)		Low	0	2	0	0.25	0.25	2	1	56	56	0	0
7	G	Symmetrical Judges (Agreement Needed)			0	2	1	0.25	0.25	2	1	56	56	0	0
8	H	Base 1	Unstable Environment		0.6	1	0	0.25	0.25	1	1	56	56	0	0
9	I	Symmetrical Judges (Agreement not Needed)		Normal	0.6	2	0	0.25	0.25	1	1	56	56	0	0
10	J	Symmetrical Judges (Agreement Needed)			0.6	2	1	0.25	0.25	1	1	56	56	0	0
11	K	Symmetrical Judges (Agreement not Needed)		High	0.6	2	0	0.25	0.25	0.25	1	56	56	0	0
12	L	Symmetrical Judges (Agreement Needed)			0.6	2	1	0.25	0.25	0.25	1	56	56	0	0
13	M	Symmetrical Judges (Agreement not Needed)		Low	0.6	2	0	0.25	0.25	2	1	56	56	0	0
14	N	Symmetrical Judges (Agreement Needed)			0.6	2	1	0.25	0.25	2	1	56	56	0	0
15	O	Asymmetrical Judges (Agreement not Needed)	Perfect Environment	Normal	0	2	0	0.25	0.5	1	1	56	25	0	0
16	P	Asymmetrical Judges (Agreement Needed)			0	2	1	0.25	0.5	1	1	56	25	0	0
17	Q	Asymmetrical Judges (Agreement not Needed)		High	0	2	0	0.25	0.5	0.25	1	56	25	0	0
18	R	Asymmetrical Judges (Agreement Needed)			0	2	1	0.25	0.5	0.25	1	56	25	0	0

19	S	Asymmetrical Judges (Agreement not Needed)		Low	0	2	0	0.25	0.5	2	1	56	25	0	0
20	T	Asymmetrical Judges (Agreement Needed)			0	2	1	0.25	0.5	2	1	56	25	0	0
21	U	Asymmetrical Judges (Agreement not Needed)	Unstable Environment	Normal	0.6	2	0	0.25	0.5	1	1	56	25	0	0
22	V	Asymmetrical Judges (Agreement Needed)			0.6	2	1	0.25	0.5	1	1	56	25	0	0
23	W	Asymmetrical Judges (Agreement not Needed)		High	0.6	2	0	0.25	0.5	0.25	1	56	25	0	0
24	X	Asymmetrical Judges (Agreement Needed)			0.6	2	1	0.25	0.5	0.25	1	56	25	0	0
25	Y	Asymmetrical Judges (Agreement not Needed)		Low	0.6	2	0	0.25	0.5	2	1	56	25	0	0
26	Z	Asymmetrical Judges (Agreement Needed)			0.6	2	1	0.25	0.5	2	1	56	25	0	0
27	AA	Symmetrical Judges (Using other Judgment)	Perfect Environment	Normal	0	2	0	0.25	0.25	1	1	56	56	0.5	0.5
28	BB	Symmetrical Judges (Using other Judgment)			0	2	1	0.25	0.25	1	1	56	56	0.5	0.5
29	CC	Asymmetrical Judges (Using other Judgment)			0	2	0	0.25	0.5	1	1	56	25	0.25	0.5
30	DD	Asymmetrical Judges (Using other Judgment)			0	2	1	0.25	0.5	1	1	56	25	0.25	0.5
31	EE	Symmetrical Judges (Using other Judgment)	Unstable	Normal	0.6	2	0	0.25	0.25	1	1	56	56	0.5	0.5
32	FF	Symmetrical Judges (Using other Judgment)			0.6	2	1	0.25	0.25	1	1	56	56	0.5	0.5
33	GG	Asymmetrical Judges (Using other Judgment)			0.6	2	0	0.25	0.5	1	1	56	25	0.25	0.5
34	HH	Asymmetrical Judges (Using other Judgment)			0.6	2	1	0.25	0.5	1	1	56	25	0.25	0.5

Results and Discussion

In the model we use stochasticity to generate information cues about the environment and about judgment predictability. In order to understand more clearly the simulation results derived from the different simuation scenarios described in the previous section we simulated each scenario 2,000 times by using different seeds for the stochastic processess driving the behavior of the information cues, the predictability of the environment, and judgment reliability. In Table 2, we report results for performance metrics of the detection-selection process: the decision threshold for Judge A, decision threshold for Judge B, selection rate, error rate, sensitivity (proportion of elements belonging to the signal distribution that are acted upon), and specificity (proportion of elements belonging to the noise distribution that are correctly ignored).

Table 2—Results (2,000 simulations)

#	ID	Decision Threshold for Judge A					Decision Threshold for Judge B					Selection Rate					Error Rate				
		Min	Max	Mean	Median	StDev	Min	Max	Mean	Median	StDev	Min	Max	Mean	Median	StDev	Min	Max	Mean	Median	StDev
1	A	58	64	60.91	61	1.092	58	64	60.91	61	1.092	0.1	0.31	0.1909	0.19	0.0341	0.02	0.16	0.0693	0.07	0.0193
2	B	59	65	62.11	62	1.014	59	65	62.11	62	1.014	0.1	0.33	0.2029	0.2	0.0348	0.03	0.16	0.0742	0.07	0.0171
3	C	57	63	59.46	59	0.9774	57	63	59.46	59	0.9774	0.07	0.3	0.1763	0.17	0.0342	0.01	0.12	0.0533	0.05	0.0178
4	D	60	66	62.64	63	0.9626	60	66	62.64	63	0.9626	0.1017	0.2183	0.1592	0.1588	0.0172	0.0223	0.0794	0.0482	0.0471	0.0092
5	E	57	63	59.82	60	0.9514	57	63	59.82	60	0.9514	0.0967	0.2109	0.1522	0.1513	0.0173	0.0173	0.0769	0.0427	0.0421	0.0094
6	F	57	65	60.86	61	1.126	57	65	60.86	61	1.126	0.1	0.42	0.2375	0.24	0.0510	0.02	0.24	0.1018	0.1	0.0263
7	G	56	62	58.70	59	0.9936	56	62	58.70	59	0.9936	0.04	0.36	0.1944	0.2	0.0504	0	0.16	0.0646	0.06	0.0257
8	H	53	65	59.58	60	1.593	53	65	59.58	60	1.593	0.12	0.34	0.2136	0.21	0.0355	0.04	0.25	0.1320	0.13	0.0321
9	I	56	66	60.84	61	1.573	56	66	60.84	61	1.573	0.13	0.36	0.2261	0.22	0.0358	0.04	0.27	0.1356	0.13	0.0318
10	J	53	63	58.17	58	1.563	53	63	58.17	58	1.563	0.09	0.33	0.1994	0.2	0.0352	0.04	0.24	0.1239	0.12	0.0321
11	K	56	67	61.21	61	1.556	56	67	61.21	61	1.556	0.1339	0.2506	0.1905	0.1910	0.0184	0.0669	0.1861	0.1228	0.1215	0.0165
12	L	54	65	58.41	58	1.520	54	65	58.41	58	1.520	0.1265	0.2431	0.1836	0.1836	0.0184	0.0669	0.1712	0.1196	0.1191	0.0166
13	M	55	65	59.92	60	1.512	55	65	59.92	60	1.512	0.08	0.42	0.2541	0.26	0.0528	0.02	0.32	0.1480	0.14	0.0452
14	N	53	63	57.75	58	1.490	53	63	57.75	58	1.490	0.04	0.38	0.2106	0.2	0.0511	0	0.3	0.1272	0.12	0.0462
15	O	86	95	90.69	91	1.495	55	64	59.69	60	1.495	0.36	0.59	0.4887	0.49	0.0346	0.3	0.43	0.3605	0.36	0.0194
16	P	58	64	60.91	61	1.092	27	33	29.91	30	1.092	0.1	0.31	0.1909	0.19	0.0341	0.02	0.16	0.0693	0.07	0.0193
17	Q	89	98	93.29	93	1.477	58	67	62.29	62	1.477	0.1786	0.2903	0.2353	0.2357	0.0170	0.1166	0.2034	0.1544	0.1538	0.0128
18	R	58	64	61.33	61	1.065	27	33	30.33	30	1.065	0.0992	0.2158	0.1560	0.1563	0.0171	0.0198	0.0967	0.0520	0.0521	0.0104
19	S	77	90	84.45	84	1.815	46	59	53.45	53	1.815	0.52	0.88	0.7093	0.7	0.0471	0.42	0.68	0.5698	0.56	0.0365
20	T	56	63	59.82	60	1.124	25	32	28.82	29	1.124	0.06	0.4	0.2167	0.22	0.0508	0	0.22	0.0868	0.08	0.3276
21	U	84	95	89.53	90	1.688	53	64	58.53	59	1.688	0.39	0.65	0.5131	0.51	0.0362	0.3	0.49	0.3777	0.38	0.0265
22	V	53	65	59.58	60	1.593	22	34	28.58	29	1.593	0.12	0.34	0.2136	0.21	0.0355	0.04	0.25	0.1320	0.13	0.2435
23	W	86	98	91.71	92	1.731	55	67	60.71	61	1.731	0.2059	0.3275	0.2662	0.2655	0.0183	0.1464	0.2704	0.2013	0.2009	0.0169
24	X	55	65	59.90	60	1.580	24	34	28.90	29	1.580	0.1290	0.2481	0.1873	0.1885	0.0184	0.0694	0.1786	0.1238	0.1240	0.0169
25	Y	77	91	83.50	83.5	1.838	46	60	52.50	52.5	1.838	0.54	0.88	0.7258	0.72	0.0491	0.42	0.7	0.5582	0.56	0.0390
26	Z	54	64	58.87	59	1.515	23	33	27.87	28	1.515	0.06	0.4	0.2330	0.24	0.0523	0.02	0.3	0.1396	0.14	0.0462
27	AA	58	64	60.72	61	0.9198	58	64	60.72	61	0.9198	0.09	0.31	0.1890	0.19	0.0345	0.02	0.15	0.0600	0.06	0.0162
28	BB	58	64	60.72	61	0.9198	58	64	60.72	61	0.9198	0.09	0.31	0.1890	0.19	0.0345	0.02	0.15	0.0600	0.06	0.0162
29	CC	86	94	90.29	90	1.306	55	63	59.29	59	1.306	0.37	0.59	0.4846	0.48	0.0346	0.3	0.4	0.3470	0.35	0.0153
30	DD	57	64	60.81	61	1.029	26	33	29.81	30	1.029	0.08	0.31	0.1899	0.19	0.0346	0.02	0.15	0.0660	0.06	0.0183
31	EE	55	64	59.45	59	1.556	55	64	59.45	59	1.556	0.11	0.34	0.2123	0.21	0.0356	0.02	0.26	0.1273	0.13	0.0321
32	FF	55	64	59.45	59	1.556	55	64	59.45	59	1.556	0.11	0.34	0.2123	0.21	0.0356	0.02	0.26	0.1273	0.13	0.0321
33	GG	85	94	89.18	89	1.569	54	63	58.18	58	1.569	0.4	0.64	0.5096	0.51	0.0364	0.3	0.46	0.3649	0.36	0.0241
34	HH	54	65	59.53	60	1.583	23	34	28.53	29	1.583	0.11	0.34	0.2130	0.21	0.0353	0.02	0.26	0.1304	0.13	0.0319

Table 2—Results (2,000 simulations) (Cont.)

#	ID	Sensitivity					Specificity				
		Min	Max	Mean	Median	StDev	Min	Max	Mean	Median	StDev
1	A	0.5	1	0.9277	0.9333	0.0700	0.8674	0.9764	0.9308	0.9310	0.0147
2	B	0.6666	1	0.9530	1	0.0572	0.8554	0.9659	0.9210	0.9204	0.0139
3	C	0.625	1	0.9329	0.9375	0.0658	0.8915	0.9882	0.9486	0.9512	0.0132
4	D	0.7727	0.9841	0.8878	0.8905	0.0341	0.9416	0.9776	0.9622	0.9624	0.0057
5	E	0.7647	0.9791	0.8825	0.8846	0.0342	0.9461	0.9857	0.9694	0.9701	0.0058
6	F	0	1	0.9822	1	0.0593	0.7560	0.9756	0.8839	0.8837	0.0275
7	G	0	1	0.9607	1	0.0810	0.8461	1	0.9307	0.9302	0.0243
8	H	0.3	1	0.7267	0.7333	0.0950	0.8133	0.9651	0.8975	0.8987	0.0232
9	I	0.3	1	0.7524	0.75	0.0946	0.7866	0.9545	0.8877	0.8888	0.0232
10	J	0.3076	1	0.7096	0.7142	0.0958	0.8266	0.9764	0.9110	0.9125	0.0225
11	K	0.5172	0.8387	0.6891	0.6901	0.0463	0.875	0.9507	0.9173	0.9176	0.0109
12	L	0.5217	0.8333	0.6786	0.6805	0.0463	0.8871	0.9554	0.9234	0.9237	0.0109
13	M	0	1	0.8000	0.8181	0.1352	0.7250	0.9772	0.8619	0.8607	0.0381
14	N	0	1	0.7340	0.75	0.1458	0.775	1	0.9010	0.9024	0.0359
15	O	0.6666	1	0.9513	0.9523	0.0572	0.5124	0.6666	0.5872	0.5882	0.0229
16	P	0.5	1	0.9277	0.9333	0.0700	0.8674	0.9764	0.9308	0.9310	0.0158
17	Q	0.6078	0.9365	0.7820	0.7844	0.0461	0.8230	0.8815	0.8558	0.8563	0.0088
18	R	0.6888	0.9848	0.8630	0.8644	0.0388	0.9341	0.9802	0.9618	0.9623	0.0063
19	S	0	1	0.9965	1	0.0315	0.1764	0.4897	0.3364	0.3404	0.0437
20	T	0	1	0.9616	1	0.0804	0.7906	1	0.9048	0.9069	0.0277
21	U	0.5384	1	0.8787	0.8846	0.0758	0.4675	0.6559	0.5656	0.5662	0.0273
22	V	0.3	1	0.7267	0.7333	0.0950	0.8133	0.9651	0.8975	0.8987	0.0232
23	W	0.5	0.8484	0.6813	0.6849	0.0477	0.7692	0.8629	0.8235	0.8238	0.0123
24	X	0.5	0.8125	0.6774	0.6794	0.0469	0.8773	0.9548	0.9187	0.9189	0.0112
25	Y	0	1	0.9752	1	0.0617	0.1351	0.4680	0.3260	0.3255	0.0471
26	Z	0	1	0.7634	0.7777	0.1414	0.75	0.9787	0.8798	0.8809	0.0375
27	AA	0.625	1	0.9539	1	0.0568	0.8674	0.9767	0.9374	0.9390	0.0127
28	BB	0.625	1	0.9539	1	0.0568	0.8674	0.9767	0.9374	0.9390	0.0127
29	CC	0.8	1	0.9852	1	0.0331	0.5256	0.6593	0.5974	0.5977	0.0209
30	DD	0.6	1	0.9356	0.9393	0.0663	0.8674	0.9767	0.9333	0.9325	0.0142
31	EE	0.2	1	0.7364	0.7391	0.0966	0.8	0.9764	0.9012	0.9024	0.0228
32	FF	0.2	1	0.7364	0.7391	0.0966	0.8	0.9764	0.9012	0.9024	0.0228
33	GG	0.6	1	0.9052	0.9130	0.0680	0.48	0.6521	0.5755	0.5764	0.0260
34	HH	0.2	1	0.7299	0.7368	0.0958	0.8	0.9764	0.8988	0.9012	0.0228

We report all characteristics of the distribution of results for these variables (minimum, maximum, mean, median, and standard deviation) and discuss the results. The different metrics chosen to capture performance will be discussed. Figure 6 shows the results obtained for selection rates and error rates across all simulation scenarios.

Runs A through G, O through T, and AA through DD (all in shadowed boxes), belong to the case of the perfectly predictable environment. It is intersting to see that selection rates are not very different in the first block of simulations (A–G) with the perfect environment than in the second block (H–N) with the unstable environment.

An exception is noted in the case of runs F and M, where a spike in selection occurs in the case of low frequency of learning and no coordination needed. Error rates, however, in the same two initial blocks vary dramatically (three fold). This result confirms our thinking that under conditions of perfect predictability of the environment, lower error rates, and better performance will be achieved (proposition 1). Higher selection rates, however, can have a negative impact on performance as this represents additional effort and cost. When asymmetrical judgment is introduced (blocks O–P and U–Z), it is quite evident that both rates experience significant increases and change dramatically when coordination of action is present. Also, it is interesting to see that in runs T and V, the standard deviation of the error rate changes significantly from all other runs. An explanation for this result is not evident to us at this time.

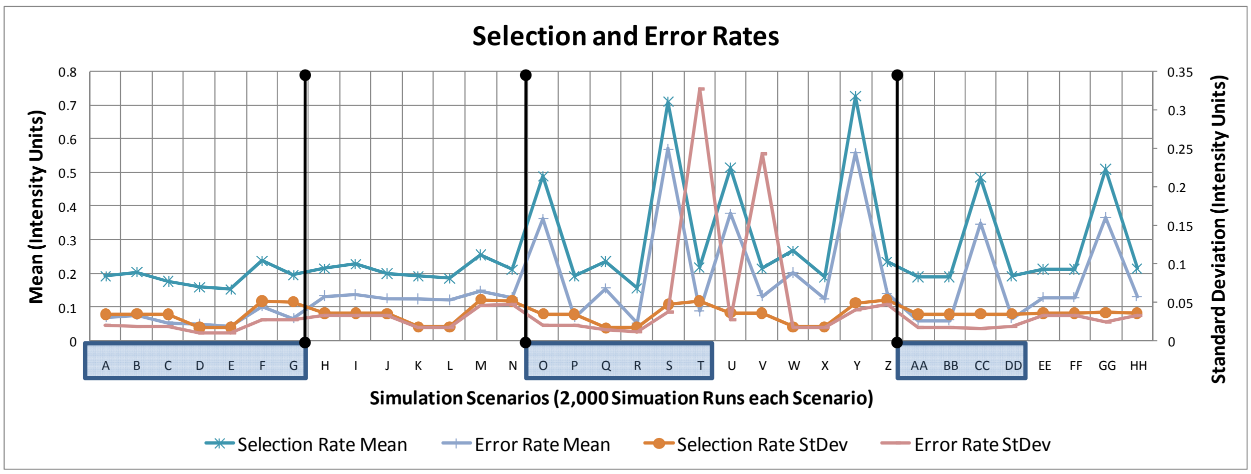
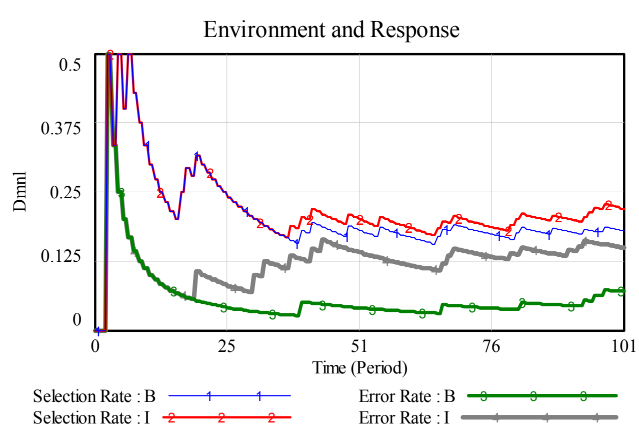


Figure 6—Selection and Error Rates Results

As presented in proposition 2, the frequency of learning influences performance. Selection rates and error rates change with the different levels of frequency of feedback. In Figure 7, we show the impact of the different levels of frequency on the temporal dynamics of selection rates and error rates. Noteworthy is the fact that the lower the frequency of feedback, the higher the variability in rates experienced (see Figure 7, lower graphs). In the graph presented in Figure 7 (lower right), a very tight behavior (compared to the others) is observed. This graph (runs D and K) captures the high-frequency case. This result was not anticipated by the authors.



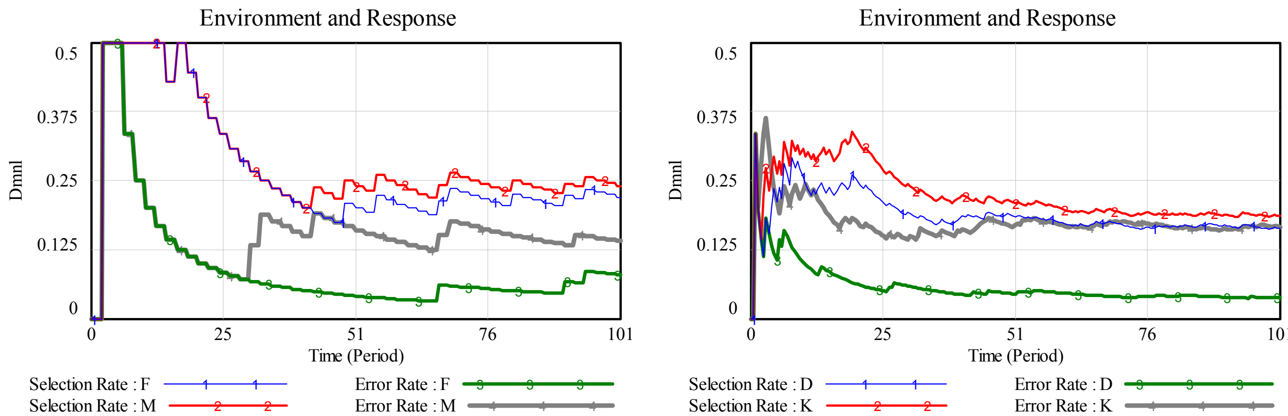


Figure 7—Effect of Frequency of Learning on Performance

The decision threshold of the judges is another good metric of performance of the process. If the judges can learn about the distal variable and find the decision threshold that minimizes error, the detection-selection process will yield a minimum error rate at the most acceptable cost. In Figure 8, we present the results for the decision thresholds of Judges A and B. We also present results of variability of the results obtained. As shown in Figure 8, in runs A through N, both judges have identical decision thresholds as these runs are parametrized for symmetrical judges. In runs O through Z, asymmetrical judges are simulated. The results follow a particular trend by showing the influence of coordination in decision making. The decision thresholds show a significant change in level when coordination is present. Also, under conditions of an unstable environment, it is noteworthy to acknowledge the significant increase in variability in the determination of the decision threshold (see standard deviation of runs U–Z as compared to that of runs O–T). This effect is also evident when comparing runs AA–DD (perfect environment) with runs EE–HH (unstable environment). The fact that the decision thresholds of both judges separate from run O forward is a function of the parametrization of the asymmetrical-judges condition. Additional research is needed to explore the conditions under which, after a definitional split, convergence may occur (several additional feedback mechanisms might be needed for this result to occur).

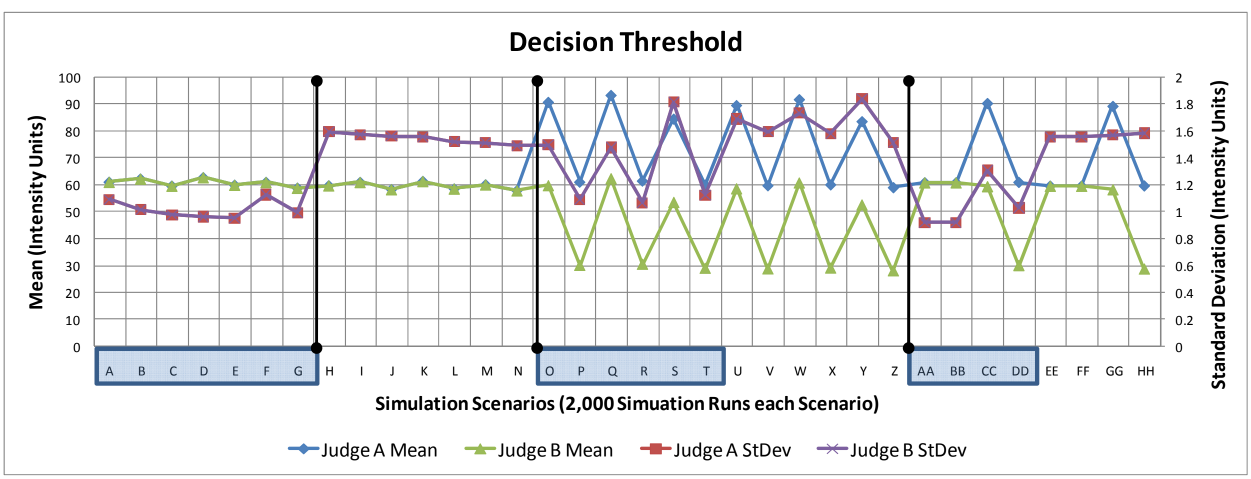


Figure 8—Decision Threshold Results

In detection-selection processes, two metrics are used to capture performance in a general way: sensitivity and specificity. Sensitivity refers to the ability of the judge to correctly identify members of the signal distribution as a proportion of the total membership. The equation for sensitivity is of the form:

$$Sensitivity_i = \sum_{i=1}^{n} \frac{TP_i}{\left(TP_i + FN_i\right)}$$

where TP represents true-positive outcomes; FN represents false-negative outcomes; and n is the number of cases examined. When sensitivity is 1, no false-negative outcomes are generated, and 100% of the members of the signal distribution are correctly identified.

Specificity refers to the ability of the judge (or system) to correctly identify members of the noise distribution as a percentage of the total membership. The equation for sensitivity is of the form:

$$Specificity_i = \sum_{i=1}^{n} \frac{TN_i}{\left(TN_i + FP_i\right)}$$

where TN represents true-negative outcomes; FP represents false-positive outcomes; and n is the number of cases examined. When specificity is 1, no false-positive outcomes are generated, and 100% of the members of the noise distribution are correctly identified.

In Figure 9, we show the results for sensitivity and specificity for all scenarios simulated. As shown in Figure 9, the sensitivity metric changes much more than the specificity metric as changes to environmental predictability are introduced (see the difference in behavior between the blocks of runs A–G and H–N). An important drop in sensitivity is clear in block H–N when compared to block A–G. Of note is the fact that low frequency of feedback (runs F–G and M–N) seem to improve the sensitivity metric while, at the same time, degrade the specificity metric independently of environmental predictability. Also, variability of the sensitivity metrics is higher (and much more volatile) than that of the specificity metric across all scenario runs. This result is interesting in that the sensitivity metric is the one that captures elements related to the signal distribution that, potentially and depending on the context, are much more costly and valuable than members of the noise distribution (characteristic captures with the specific metric). This result is important because under conditions of high uncertainty and low base rates (as in the case of the identification of terrorists), small changes in detection of members of the signal distribution (i.e., the population of terrorists) will have a very important effect on the performance of the detection-selection system. Alternately, high levels of specificity are indicative of the ability to avoid unnecessary inconveniences to members of the noise distribution (non-terrorists distribution). This result, although very desirable and potentially important, might not be crucial to the mission of identification of members of the signal distribution.

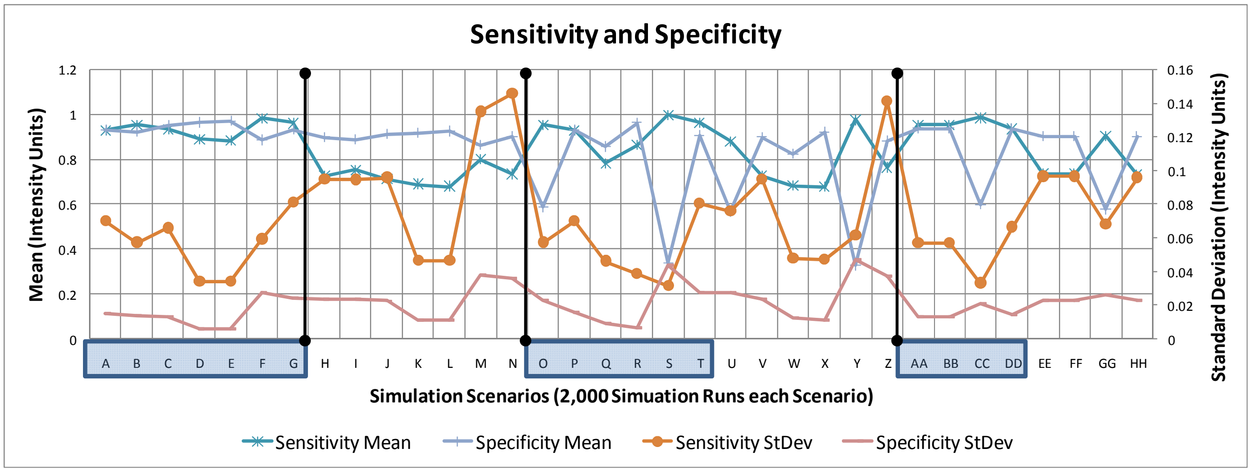


Figure 9—Sensitivity and Specificity Results

The 34 scenarios simulated provide a very rich source of information to further analyze how performance of detection-selection systems can be improved. In Appendix 2, we compare and contrast different sets of runs and scenario characterizations that will be analyzed in future research.

The dual-judge model allowed us to identify the finding that, under conditions of high uncertainty, having a dual-judge model yields better results than found in case of using a single-judge and captures in a better way the conditions that human judges face in detection-selection processes. However, the dual-judge process, under conditions in which there is no coordination, has the potential to perform extremely poorly and with high levels of variability. In multi-judge environments, it appears that coordination of action is very useful in achieving better results. Also, contrary to our expectations for the findings, higher levels of frequency of feedback might not generate only good results. Additional research into these results is needed.

Acknowledgements

This work was funded in part by the U.S. Department of Homeland Security.

References

Brunswik, E. 1943. Organismic achievement and environmental probability. Psychological Review, 50: 255-272.

Brunswik, E. 1956. Perception and the Representative Design of Psychological Experiments. Berkeley, CA: University of California Press.

Egan, J. P. 1975. Signal Detection Theory and ROC Analysis. New York: Academic Press.

Erev, I., Gopher, D., Itkin, R., & Greenshpan, Y. 1995. Toward a Generalization of Signal Detection Theory to N-Person Games: The Example of Two-Person Safety Problem. Journal of Mathematical Psychology, 39: 360-375.

Erev, I. 1998. Signal detection by human observers: A cutoff reinforcement learning model of categorization decisions under uncertainty. Psychological Review, 105: 280-298.

Forrester, J. W. 1961. Industrial Dynamics. In C. Heyel (Ed.), The Encyclopedia of Management: 313-319. New York: Reinhold Publishing Company.

Ghaffarzadegan, N. 2008. Effect of Conditional Feedback on Learning, System Dynamics Conference. Athens, Greece.

Green, D. M. & Swets, J. A. 1966. Signal Detection Theory and Psychophysics. New York: John Wiley.

Hammond, K. R. 1996. Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice. New York: Oxford University Press.

Hammond, K. R. 2000. Judgments Under Stress. New York: Oxford University Press.

Hammond, K. R. & Stewart, T. R. (Eds.). 2001. The Essential Brunswik: Beginnings, Explications, Applications. New York: Oxford University Press.

Macmillan, N. A. & Creelman, C. D. 2005. Detection Theory: A User's Guide. New York: Lawrence Erlbaum Associates.

Martinez-Moyano, I. J., Rich, E. H., Conrad, S. H., Stewart, T., & Andersen, D. F. 2006. Integrating Judgment and Outcome Decomposition: Exploring Outcome-based Learning Dynamics, International Conference of the System Dynamics Society. Nijmegen, The Netherlands: System Dynamics Society.

Martinez-Moyano, I. J., Conrad, S. H., & Andersen, D. F. 2007. An Outcome-Based Learning Model to Identify Emerging Threats: Experimental and Simulation Results. Paper presented at the 40th Annual Hawaii International Conference on System Sciences (HICSS'07), Hawaii, HI.

Martinez-Moyano, I. J., Rich, E., Conrad, S., Andersen, D. F., & Stewart, T. R. 2008. A behavioral theory of insider-threat risks: A system dynamics approach. ACM Trans. Model. Comput. Simul., 18(2): 1-27.

Richardson, G. P. & Pugh, A. L., III. 1981. Introduction to System Dynamics Modeling. Waltham MA: Pegasus Communications, Inc.

Sterman, J. D. 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World. Boston MA: Irwin McGraw-Hill.

Swets, J. 1992. The Science of Choosing the Right Decision Threshold in High-Stakes Diagnostics. American Psychologist, 47(4): 522-532.

Weaver, E. & Richardson, G. P. 2002. Threshold Setting and the Cycling of a Decision Threshold. System Dynamics Review, 22(1): 1-26.

Appendix 1—Characterization of Judges and Model Flags

Table A—Judge A Parameters

Variable Name	Description	Default Value	Units
Decision Threshold Initial Judge A	This parameter captures the initial condition of the decision threshold for Judge A. The range of this variable is between 0 and 100. A threshold of 0 represents a super-vigilant judge that will scrutinize everything, while a threshold of 100 represents a super-trusting judge that will not trigger any action ever.	56	Intensity
Judge A Weight of Judgment of Judge B	This parameter captures how important the judgment of Judge B is for Judge A's judgment. This formulation generates a double judgment process in which first the judge uses environmental information cues to make a judgment about the distal variable and then, in a second judgment, incorporates information about the other judge's judgment. When this weight is 0, the judge is not using the other judge's judgment; when this weight is 1; Judge A is using Judge B's judgment to make his judgment (disregarding his own judgment).	0	Dmnl
Influence of False Negatives on Decision Threshold Judge A	This parameter captures the value that information about a false-negative outcome has on the way in which Judge A adjusts his/her decision threshold. When the value is zero, no adjustment is made to the threshold. When the value is positive, the threshold is adjusted up (making the judge's behavior more trusting). When the value is negative, the threshold is adjusted down (making the judge's behavior more vigilant).	-1	Dmnl/Event
Influence of False Positives on Decision Threshold Judge A	This parameter captures the value that information about a false-positive outcome has on the way in which Judge A adjusts his/her decision threshold. When the value is zero, no adjustment is made to the threshold. When the value is positive, the threshold is adjusted up (making the judge's behavior more trusting). When the value is negative, the threshold is adjusted down (making the judge's behavior more vigilant).	1	Dmnl/Event
Influence of True Positives on Decision Threshold Judge A	This parameter captures the value that information about a true-positive outcome has on the way in which Judge A adjusts his/her decision threshold. When the value is zero, no adjustment is made to the threshold. When the value is positive, the threshold is adjusted up (making the judge's behavior more trusting). When the value is negative, the threshold is adjusted down (making the judge's behavior more vigilant).	0	Dmnl/Event
Influence of True Negatives on Decision Threshold Judge A	This parameter captures the value that information about a true-negative outcome has on the way in which Judge A adjusts his/her decision threshold. When the value is zero, no adjustment is made to the threshold. When the value is positive, the threshold is adjusted up (making the judge's behavior more trusting). When the value is negative, the threshold is adjusted down (making the judge's behavior more vigilant).	0	Dmnl/Event
Judge A Weight of Information Cue 1	This parameter captures how important it is for Judge A Information Cue 1 in the conformation of the judgment of the distal variable. When this weight is 0, Information Cue 1 is not used by Judge A. The weights in this model are normalized to 1.	(1/6)	Dmnl
Judge A Weight of Information Cue 2	This parameter captures how important it is for Judge A Information Cue 2 in the conformation of the judgment of the distal variable. When this weight is 0, Information Cue 2 is not used by Judge A. The weights in this model are normalized to 1.	(2/6)	Dmnl
Judge A Weight of Information Cue 3	This parameter captures how important it is for Judge A Information Cue 3 in the conformation of the judgment of the distal variable. When this weight is 0, Information Cue 3 is not used by Judge A. The weights in this model are normalized to 1.	(3/6)	Dmnl
Judge A Bias	This parameter captures a bias in judgment for Judge A. In theory, bias can range between -100 and +100. When this variable is 0, it represents an unbiased judge. The judgment of likelihood is a variable that ranges from 0 to 100, and as the bias component has an additive effect on judgment (capped at the frontiers of its range), the bias parameter can make the judgment extreme in the range independently of the other components of judgment. This parameter captures the possibility of having extremely biased judges that, independent of the evidence provided to them, will judge the likelihood of a phenomenon according to their prior beliefs.	0	Intensity
Judge A Reliability Weight	This parameter captures how reliable Judge A is. A zero (0) weight represents a perfectly reliable judge as it cancels a stream of stochastic noise introduced to the judgment process (a judge that assesses the evidence in a perfectly consistent way over time). A weight of 1 represents a judge that is as unstable as it can possible be given a certain distribution of stochastic noise in its reliability of judgment (a weight of 1 lets through 100% of the reliability noise into the judgment process). The stochastic characterization of the reliability noise is captured elsewhere.	0.1	Dmnl

Judge A Reliability Error StdD	This parameter captures the standard deviation of Judge A's reliability noise stream. The judgment of the distal variable varies between 0 and 100. Therefore, the higher the standard deviation for the noise stream, the higher the uncertainty and reliability of the judgment process. This parameter captures (in part) the inherent stability and reliability of the cognitive process of the judge when making repeated judgments. When this parameter is 0, the judge is perfectly reliable (always uses the information cues in the exact same way). When this parameter is 100 (when paired with minimum and maximum of -100 and +100 in a truncated distribution and with a weight of 1 for the judge's reliability), the judge is completely erratic and unreliable in his/her judgment. Articulated differently, the judgment is random, independent of the information cues about the distal variable.	10	Intensity
Judge A Reliability Error Min	This parameter captures the minimum value for Judge A's reliability noise stream. This value truncates the results of the stochastic noise to this minimum value. The noise stream is determined to be centered on 0 (also a parameter). The combination of this parameter with the Error Max parameter generates the range of variability of the reliability noise stream for Judge A.	-30	Intensity
Judge A Reliability Error Max	This parameter captures the maximum value for Judge A's reliability noise stream. This value truncates the results of the stochastic noise to this maximum value. The noise stream is determined to be centered on 0 (also a parameter). The combination of this parameter with the Error Min parameter generates the range of variability of the reliability noise stream for Judge A.	30	Intensity

Table B—Judge B Parameters

Variable Name	Description	Default Value	Units
Decision Threshold Initial Judge B	This parameter captures the initial condition of the decision threshold for Judge B. The range of this variable is between 0 and 100. A threshold of 0 represents a super-vigilant judge that will scrutinize everything, while a threshold of 100 represents a super-trusting judge that will not trigger any action ever.	56	Intensity
Judge B Weight of Judgment of Judge A	This parameter captures how important the judgment of Judge A is for Judge B's judgment. This formulation generates a double judgment process in which first the judge uses environmental information cues to make a judgment about the distal variable and then, in a second judgment, incorporates information about the other judge's judgment. When this weight is 0, the judge is not using the other judge's judgment; when this weight is 1; Judge B is using Judge A's judgment to make his judgment (disregarding his own judgment).	0	Dmnl
Influence of False Negatives on Decision Threshold Judge B	This parameter captures the value that information about a false-negative outcome has on the way in which Judge B adjusts his/her decision threshold. When the value is zero, no adjustment is made to the threshold. When the value is positive, the threshold is adjusted up (making the judge's behavior more trusting). When the value is negative, the threshold is adjusted down (making the judge's behavior more vigilant).	-1	Dmnl/Event
Influence of False Positives on Decision Threshold Judge B	This parameter captures the value that information about a false-positive outcome has on the way in which Judge B adjusts his/her decision threshold. When the value is zero, no adjustment is made to the threshold. When the value is positive, the threshold is adjusted up (making the judge's behavior more trusting). When the value is negative, the threshold is adjusted down (making the judge's behavior more vigilant).	1	Dmnl/Event
Influence of True Positives on Decision Threshold Judge B	This parameter captures the value that information about a true-positive outcome has on the way in which Judge B adjusts his/her decision threshold. When the value is zero, no adjustment is made to the threshold. When the value is positive, the threshold is adjusted up (making the judge's behavior more trusting). When the value is negative, the threshold is adjusted down (making the judge's behavior more vigilant).	0	Dmnl/Event
Influence of True Negatives on Decision Threshold Judge B	This parameter captures the value that information about a true-negative outcome has on the way in which Judge B adjusts his/her decision threshold. When the value is zero, no adjustment is made to the threshold. When the value is positive, the threshold is adjusted up (making the judge's behavior more trusting). When the value is negative, the threshold is adjusted down (making the judge's behavior more vigilant).	0	Dmnl/Event
Judge B Weight of Information Cue 1	This parameter captures how important it is for Judge B Information Cue 1 in the conformation of the judgment of the distal variable. When this weight is 0, Information Cue 1 is not used by Judge B. The weights in this model are normalized to 1.	(1/6)	Dmnl
Judge B Weight of Information Cue 2	This parameter captures how important it is for Judge B Information Cue 2 in the conformation of the judgment of the distal variable. When this weight is 0, Information Cue 2 is not used by Judge B. The weights in this model are normalized to 1.	(2/6)	Dmnl
Judge B Weight of Information Cue 3	This parameter captures how important it is for Judge B Information Cue 3 in the conformation of the judgment of the distal variable. When this weight is 0, Information Cue 3 is not used by Judge B. The weights in this model are normalized to 1.	(3/6)	Dmnl
Judge B Bias	This parameter captures a bias in judgment for Judge B. In theory, bias can range between -100 and +100. When this variable is 0, it represents an unbiased judge. The judgment of likelihood is a variable that ranges from 0 to 100, and as the bias component has an additive effect on judgment (capped at the frontiers of its range), the bias parameter can make the judgment extreme in the range independently of the other components of judgment. This parameter captures the possibility of having extremely biased judges that, independent of the evidence provided to them, will judge the likelihood of a phenomenon according to their prior beliefs.	0	Intensity
Judge B Reliability Weight	This parameter captures how reliable Judge B is. A zero (0) weight represents a perfectly reliable judge as it cancels a stream of stochastic noise introduced to the judgment process (a judge that assesses the evidence in a perfectly consistent way over time). A weight of 1 represents a judge that is as unstable as it can possible be, given a certain distribution of stochastic noise in its reliability of judgment (a weight of 1 lets through 100% of the reliability noise into the judgment process). The stochastic characterization of the reliability noise is captured elsewhere.	0.1	Dmnl

Judge B Reliability Error StdD	This parameter captures the standard deviation of Judge B's reliability noise stream. The judgment of the distal variable varies between 0 and 100. Therefore, the higher the standard deviation for the noise stream, the higher the uncertainty and reliability of the judgment process. This parameter captures (in part) the inherent stability and reliability of the cognitive process of the judge when making repeated judgments. When this parameter is 0, the judge is perfectly reliable (always uses the information cues in the exact same way). When this parameter is 100 (when paired with minimum and maximum of -100 and +100 in a truncated distribution and with a weight of 1 for the judge's reliability), the judge is completely erratic and unreliable in his/her judgment. Articulated differently, the judgment is random, independent of the information cues about the distal variable.	10	Intensity
Judge B Reliability Error Min	This parameter captures the minimum value for Judge B's reliability noise stream. This value truncates the results of the stochastic noise to this minimum value. The noise stream is determined to be centered on 0 (also a parameter). The combination of this parameter with the Error Max parameter generates the range of variability of the reliability noise stream for Judge B.	-30	Intensity
Judge B Reliability Error Max	This parameter captures the maximum value for Judge B's reliability noise stream. This value truncates the results of the stochastic noise to this maximum value. The noise stream is determined to be centered on 0 (also a parameter). The combination of this parameter with the Error Min parameter generates the range of variability of the reliability noise stream for Judge B.	30	Intensity

Table C—Flags

Variable Name	Description	Default Value	Units
Same Size Change for Judges Flag	When this switch is on (1), the same size of change in the decision threshold jump is used for both judges to create identical behavior.	0	Dmnl
Same Judge Reliability Flag	When this switch is on (1), the same reliability is used for both judges to create identical behavior.	0	Dmnl
Same Decision Threshold Flag	If this flag is O, the judges use their own decision threshold in the decision process. When this flag is 1, Judge B uses Judge A's decision threshold in his/her decision process, disregarding his own value matrix.	0	Dmnl
Judges Involved Flag	When this flag is 1, just one judge is active (Judge A). When this flag is 2, two judges are active (A & B).	2	Dmnl
Agreement for action needed Policy Flag	This flag implements the "judges-agreement-is-needed-for-action" policy	0	Dmnl
Dynamic Threshold activation Flag for Judge A	When this flag is 0, the threshold is static (no change). When the flag is 1, the threshold is dynamic (changes as other variables change in the simulation).	1	Dmnl
Dynamic Threshold activation Flag for Judge B	When this flag is 0, the threshold is static (no change). When the flag is 1, the threshold is dynamic (changes as other variables change in the simulation).	1	Dmnl
Weight of Unpredictability of the Environment	This flag captures how inherently predictable the environment is. A zero (0) weight represents a perfectly predictable environment as it cancels a stream of stochastic noise introduced to the criterion model. A weight of 1 represents an environment that is as unstable as it can possible be, given a certain distribution of stochastic noise in its predictability. The stochastic characterization of the predictability noise is captured elsewhere (0.866025404 for 50%, 0.6 for 80% predictability, Perfect 0).	0	Dmnl

Appendix 2—Additional Graphs of Results of Runs

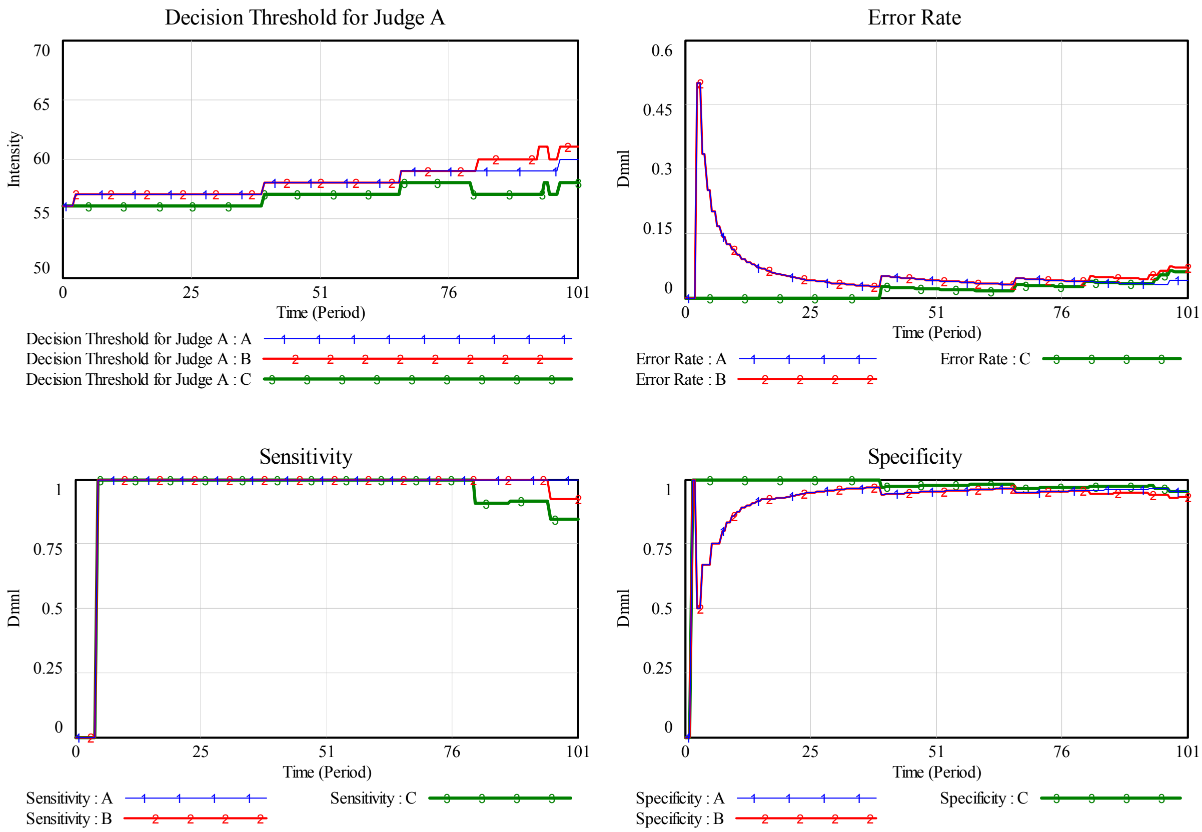
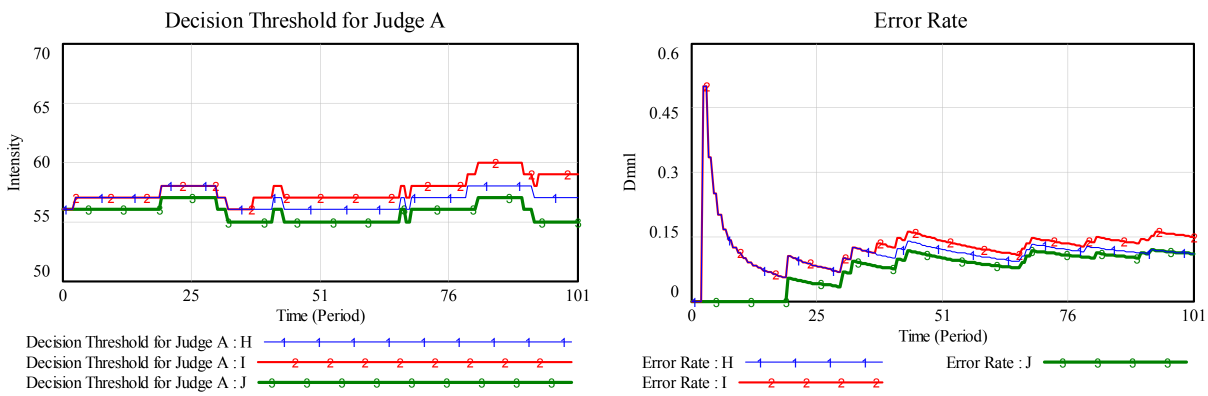


Figure 10—Symmetrical Judges Perfect Environment



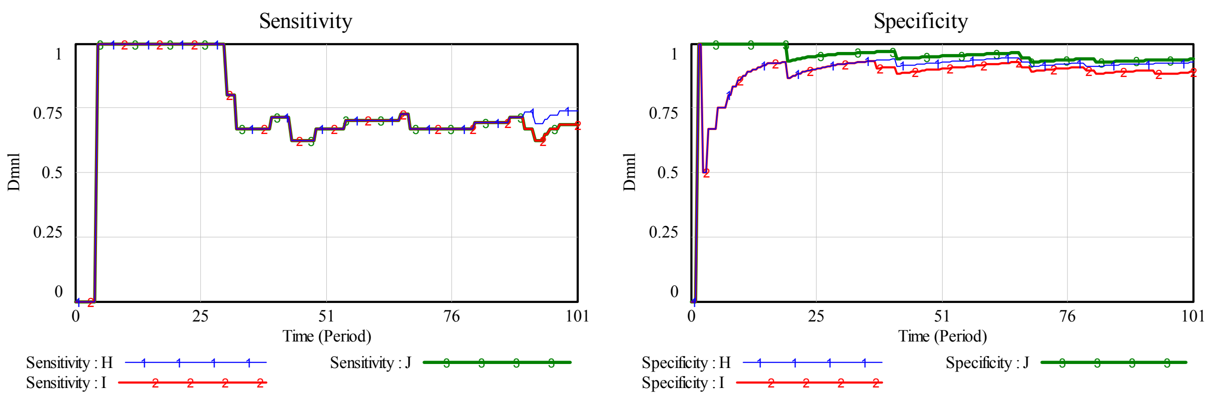


Figure 11—Symmetrical Judges Unstable Environment

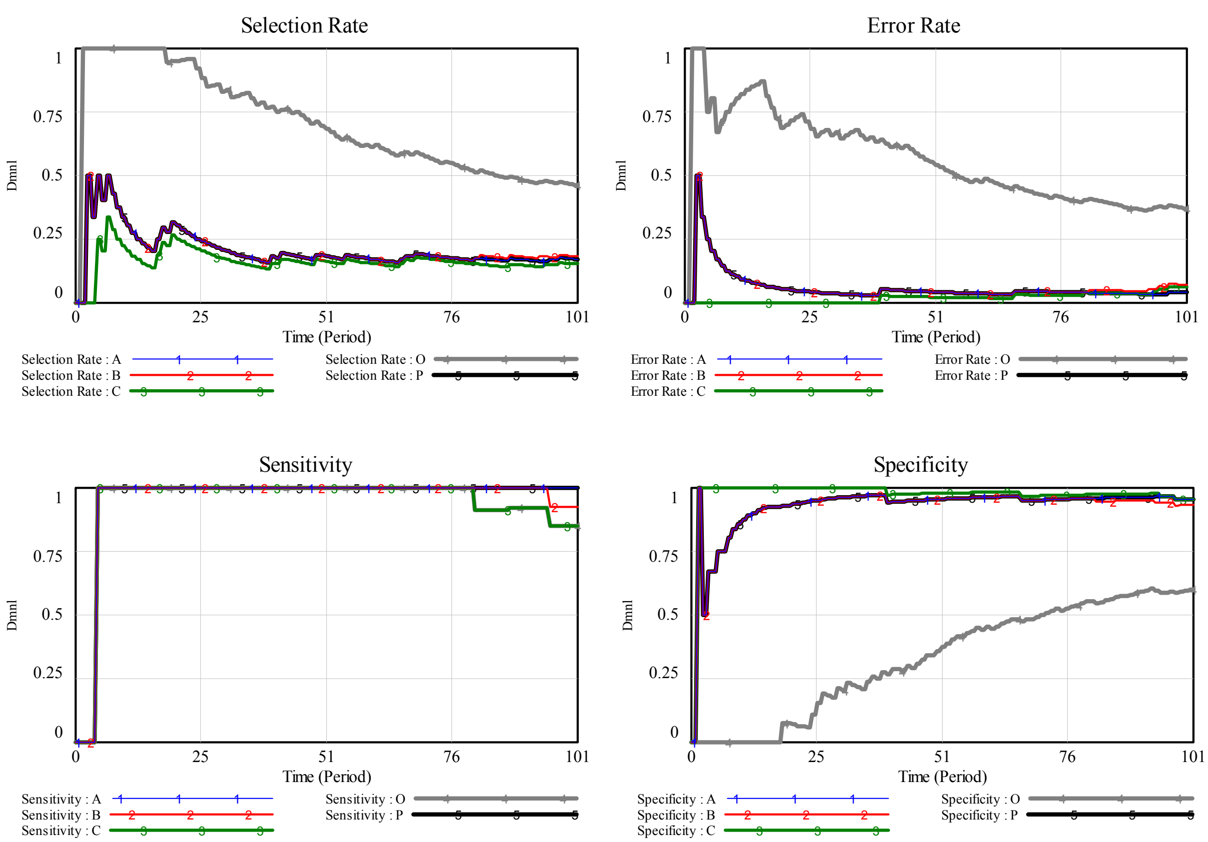


Figure 12—Asymmetrical Judges Perfect Environment

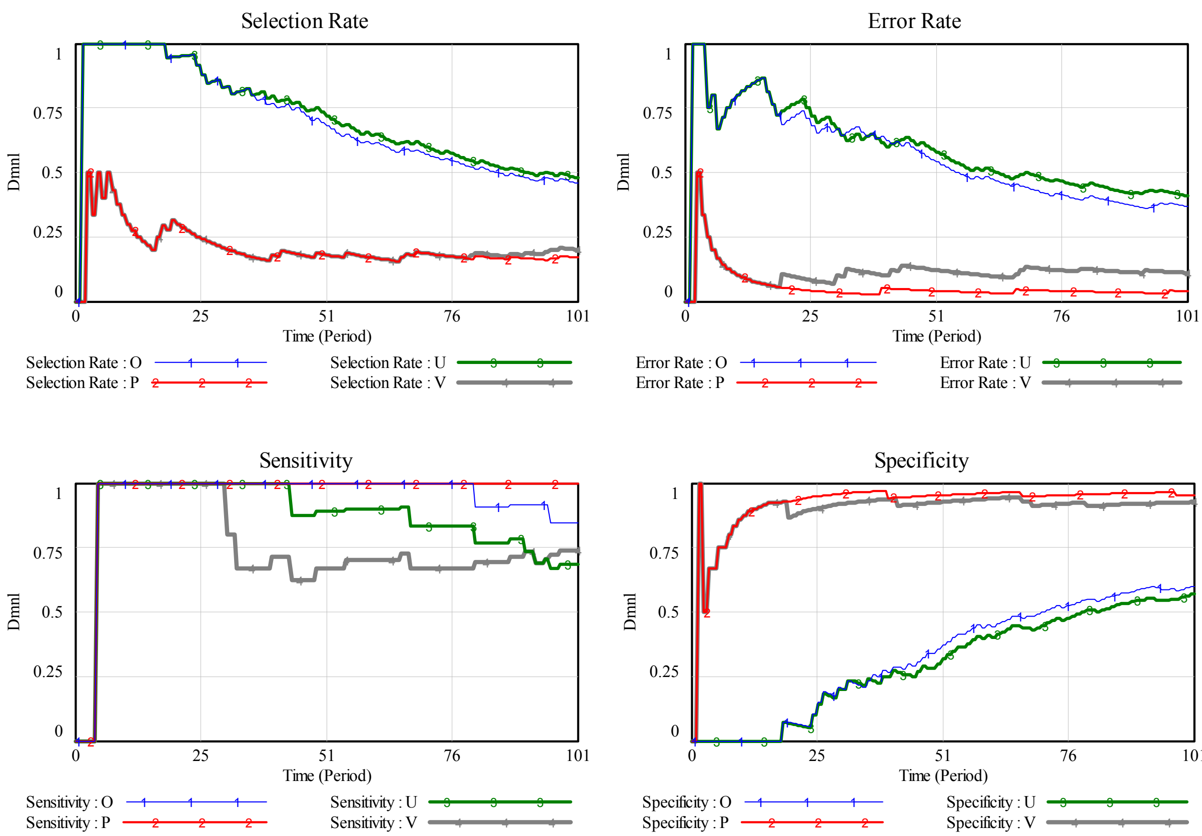


Figure 13—Asymmetrical Judges Perfect/Unstable Environment

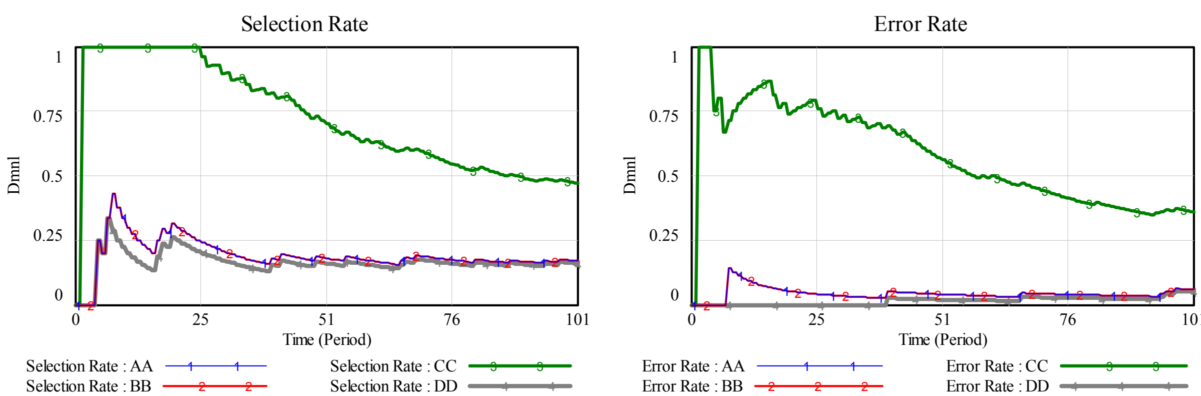


Figure 14—Symmetrical/Asymmetrical Judges Perfect Environment

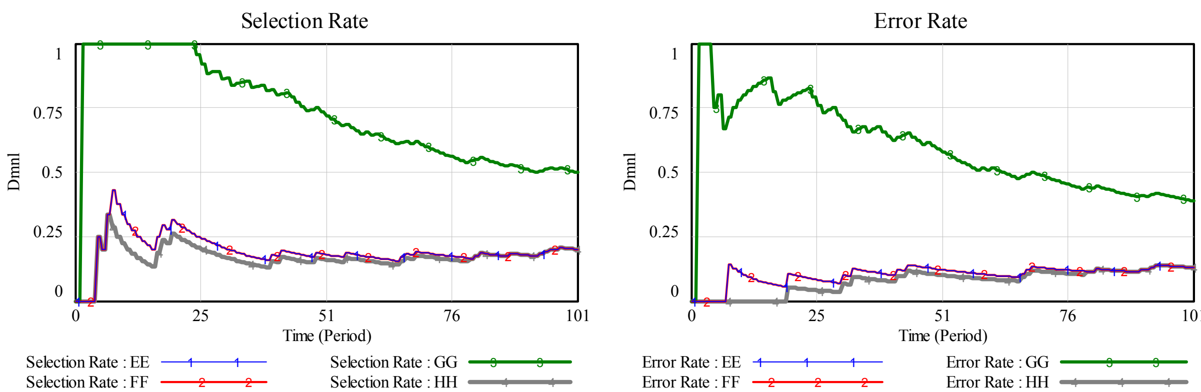


Figure 15—Symmetrical/Asymmetrical Judges Unstable Environment

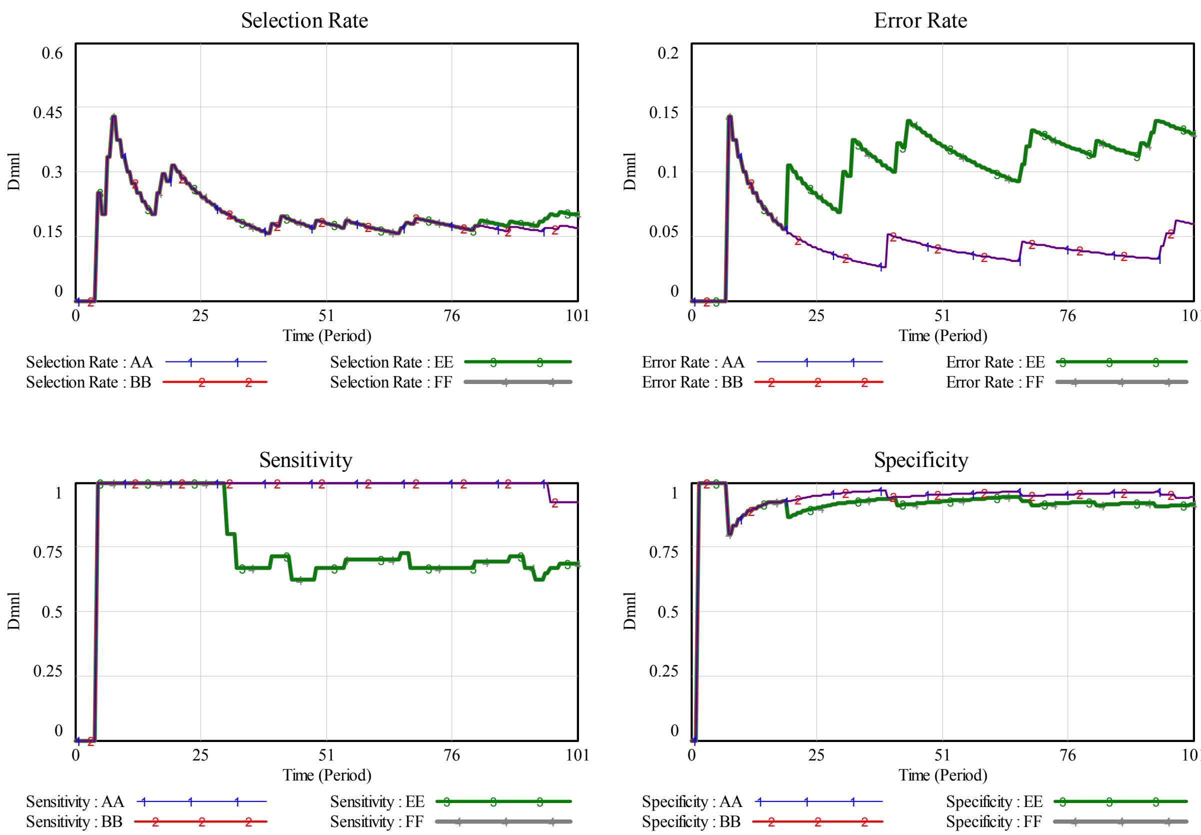


Figure 16—Symmetrical Judges with additional Information Perfect/Unstable Environment

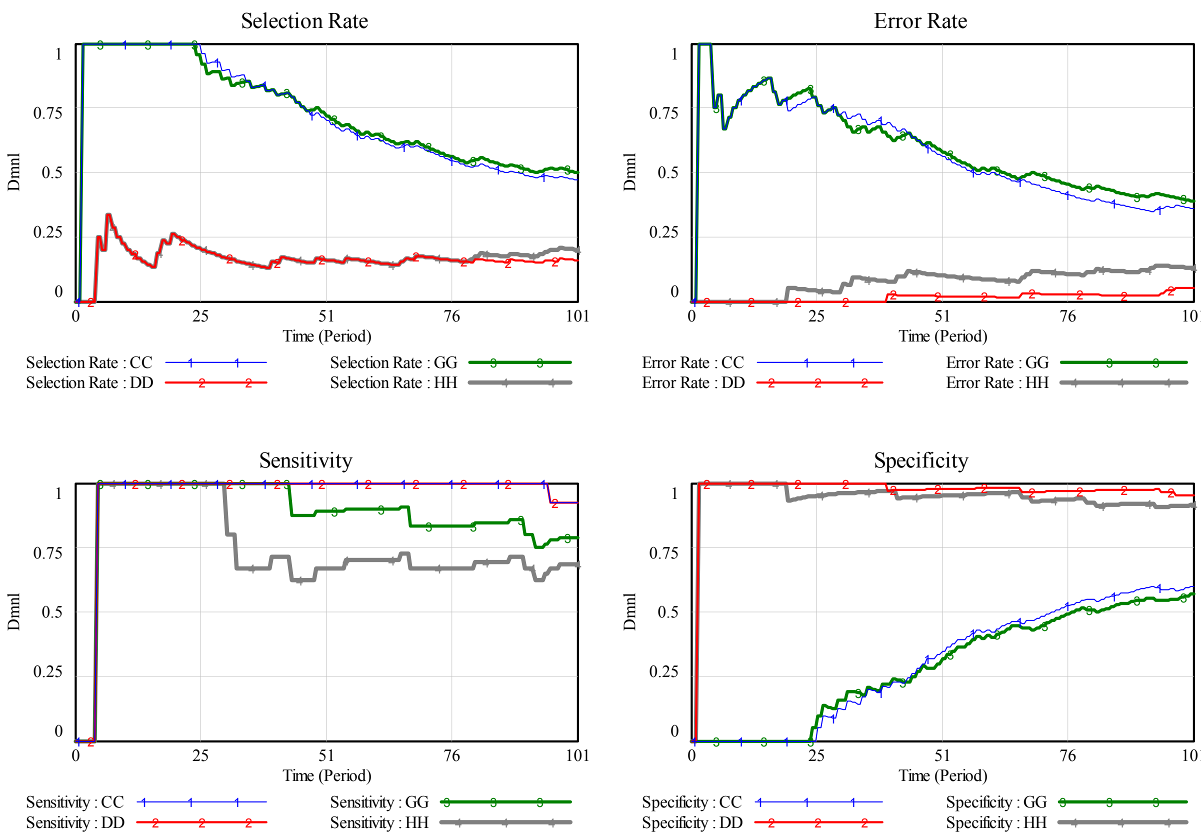


Figure 17—Asymmetrical Judges with additional Information Perfect/Unstable Environment