

Detecting malingerers by means of response-sequence analysis

MIRIAM R. LINSCHOTEN

University of Colorado Health Sciences Center, Denver, Colorado

and

LEWIS O. HARVEY, JR.

University of Colorado, Boulder, Colorado

The feigning of sensory loss (malingering) poses a challenge for psychophysicists. To uncover malingerers by means of psychophysical testing, we combined measures of response-sequence randomness with the maximum-likelihood adaptive-staircase procedure used to measure sensory detection thresholds. The two-alternative, forced-choice maximum-likelihood adaptive-staircase procedure calculates an estimate of the threshold after each trial and also recommends the stimulus concentration for the next trial. Olfactory detection thresholds for butyl alcohol were measured in 7 normals, 6 anosmics, and 6 malingerers. Each participant was tested for 20 trials. A discriminant analysis, using threshold concentration and probability of being correct over the 20 trials, could correctly classify only 68% of the malingerers and anosmics. Correct classification of anosmics and malingerers rose to 100% when statistical measures of randomness in the response sequences were included in the discriminant analysis. We conclude that the maximum-likelihood adaptive-staircase procedure, combined with response-sequence analysis, is a powerful addition to the arsenal of techniques for detecting malingerers in the evaluation of sensory ability.

Malingering is the intentional production of false or grossly exaggerated physical or psychological symptoms motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs (American Psychiatric Association, 2000). Estimates of its prevalence vary widely. In the United States, for example, Suchy and Sweet (2000) evaluated performance on the Information/Orientation (IO) subtest of the Wechsler Memory Scale-Revised (WMS-R) and estimated that 24% of benefit-seeking patients were malingering, and Andrikopoulos (2001) concluded that the malingering rate is even higher in this same population. Among Dutch patients involved in litigation after whiplash trauma,

61% were identified as malingering (Schmand et al., 1998). In Germany, 14% of 145 patients with chemosensory dysfunction after trauma or exposure to occupational and environmental toxins were classified as malingerers (Delank, Nieschalk, Schmä, & Stoll, 1999).

These high malingering rates pose serious legal and ethical problems for society. They also pose special problems for experts called upon to evaluate the sensory capacities of individuals claiming sensory and perceptual loss. In this article, we introduce a new method to better discriminate patients with a true sensory loss from malingering patients feigning a loss. We will illustrate this method in the context of chemosensory thresholds, but the described technique can be used with any sensory modality.

Broadly speaking, the general strategy for detecting malingering is to use techniques that reveal behavior inconsistent with the claimed loss. In the field of personality assessment and psychological/psychiatric disorders, for example, the use of this strategy for the detection of malingering has a long history (Hall & Poirier, 2000; Rogers, 1997). Indeed, most standard diagnostic tests include groups of items specifically designed to detect persons who are faking psychiatric disorders or who are attempting to hide psychiatric disorders (Nicholson et al., 1997). A claimed sensory loss may be the basis of a personal injury lawsuit or the seeking of disability benefits. In such cases, a detailed history of the complaint can reveal inconsistencies, but these by themselves may not be sufficient to prove malingering in a court of law.

This research was supported by NIH Grant PO1 DC00244 to M.R.L. We thank Sandra M. Mitchell for serving as the blind tester and for collecting the data. We are grateful to Naomi P. Friedman, who, as a graduate student at the University of Colorado, introduced us to the statistics of sequences. We also thank Christian Kaernbach for insightful comments that led to substantial improvement of this article. L.O.H. greatly benefited from the kind hospitality and stimulating conversations of Peter van der Helm, Emanuel Leeuwenberg, and Charles de Weert, Nijmegen Institute for Cognition and Information (NICI), University of Nijmegen, during a sabbatical visit in January and February of 2001. Some of the results reported here were first presented in 2002 at the 24th Annual Meeting of the Association for Chemoreception Sciences. Correspondence should be addressed to M. R. Linschoten, UCHSC/Box B-205, 4200 East 9th Avenue, Denver, CO 80262 (e-mail: miriam.linschoten@uchsc.edu).

Note—This article was accepted by the previous editorial team, headed by Neil Macmillan

What is needed are objective measurements of sensory performance that also can serve as indicators of deception. In vision, one can make use of the fact that it is not possible to detect which of the two eyes is being stimulated (Fahle & Mohn, 1989). In the chemical senses, Löhner (1940), assuming that olfactory information and nasal trigeminal information are conveyed in separate neural channels, used both chemosensory and trigeminal stimuli to discriminate between true and simulated anosmias. Unfortunately, with this approach, patients with a total or partial loss of olfactory sensitivity may also show a decrease in the perception of trigeminal stimuli, as measured both by event-related potentials (Hummel et al., 1996) and by psychophysical methods (Gudziol, Schubert, & Hummel, 2001). Furthermore, Leopold, Hornung, and Schwob (1992) concluded that congenital anosmic patients were unable to identify both olfactory and trigeminal odorants, although many of the patients appeared to have some—albeit diminished—ability to perceive at least some component of trigeminal stimuli.

Several suprathreshold tests of olfactory identification have the potential to discover malingerers. In the University of Pennsylvania Smell Identification Test (UPSIT), chance identification performance is 25% correct, and if one does significantly worse than that, one is classified as probably malingering (Doty, Shaman, Kimmelman, & Dann, 1984). The Odorant Confusion Matrix (OCM), a test in which 10 different odorants and a blank are each presented 10 times, uses the pattern of identification responses to detect malingerers. Malingerers described vinegar with significantly fewer irritant labels than did true anosmics (Kurtz, White, Hornung, & Belknap, 1999).

Threshold evaluation offers advantages for detecting malingerers because the limits of detection performance for normal, anosmic, and malingering observers can be predicted from detection theory (D. M. Green & Swets, 1966/1974; Macmillan & Creelman, 1991; Wickens, 2002). Figure 1 illustrates the receiver operating characteristics (ROCs) of a normal (upper-left panel), an anosmic (upper-right panel), and a malingerer with normal sensory capacity (lower-left panel) in a single-interval detection experiment using a weak stimulus. In a two-alternative, forced-choice (2AFC) paradigm, the probability of being correct corresponds to the area under the ROC curve. The normal observer obtains a hit rate (saying “yes” when the stimulus is present) that is higher than the false alarm rate (saying “yes” when the stimulus is absent). A blind observer obtains a hit rate equal to the false alarm rate. Persons who use sensory information to make the wrong response (i.e., to say “no” when the truthful answer is “yes”) will have a hit rate that is less than the false alarm rate. Their ability to give the wrong answer is limited by their ability to correctly detect the stimulus. Thus, a person with normal sensory capacity can generate hit rate–false alarm rate pairs that must lie within the response envelope shown in the lower-right panel of Figure 1. The stronger the stimulus becomes, the more bowed the ROC response envelope becomes.

To our knowledge, no threshold—as opposed to suprathreshold—tests to determine chemosensory sensitivity are currently used to detect malingerers. True anosmics would be expected to perform detection tasks at chance level. Malingerers, on the other hand, often perform at levels below chance (see, e.g., Prigatano, Smason, Lamb, & Bortz, 1997). Perhaps malingerers can be discriminated from anosmics on the basis of their detection probabilities.

In the clinic of the Rocky Mountain Taste and Smell Center, we routinely measure chemosensory thresholds using a maximum-likelihood staircase procedure, combined with a 2AFC detection paradigm (Linschoten, Harvey, Eller, & Jafek, 2001). In this method, the stimulus concentration that would give 75% correct in the 2AFC detection task is estimated after every trial on the basis of detection performance on all previous trials. The stimulus used on the next trial is the one closest to the estimated threshold concentration. A person with a normal threshold quickly converges to the threshold value, as is illustrated in the upper panel of Figure 2. An anosmic subject, lacking any ability to detect the stimulus, drives the staircase procedure to higher and higher levels of stimulus concentration (middle panel of Figure 2). A malingerer, especially one who is wrong more than predicted by chance, drives the staircase to even higher stimulus levels (lower panel of Figure 2). The final estimate of threshold concentration at the end of a relatively small number of trials will be quite high for both anosmics and malingerers and may be sufficiently different to allow detection of the malingerer.

Another possible way to detect malingerers is based on the fact that it is very difficult for people to generate truly random sequences (Bar-Hillel & Wagenaar, 1993; Brugger, 1997; Evans, 1978; Ginsburg & Karpiuk, 1994; Horne, Evans, & Orne, 1982; Neuringer, 1986; Towse & Valentine, 1997; Tune, 1964; Wagenaar, 1970, 1972). In a 2AFC paradigm, two response sequences are generated by the observer: the sequence of interval choices (1 or 2) and the sequence of correct and incorrect responses. There is a third sequence, of course: the sequence of the correct interval, which is determined by computer and normally is random. The use of measures of randomness to characterize response sequences is not new in psychology (Bar-Hillel & Wagenaar, 1993; Evans, 1978; Ginsburg & Karpiuk, 1994; Neuringer, 1986; Towse, 1998; Towse & Mclachlan, 1999; Towse & Valentine, 1997; Tulving, 1962; van der Helm, 2000; van der Helm, van Lier, & Leeuwenberg, 1992; Wagenaar, 1968, 1972), but it is new in the present context. We propose to use measures of randomness to characterize the two subject-generated sequences (right–wrong sequence and interval 1 or 2 sequence) for the purpose of detecting malingerers.

It is likely that the malingerer, who, of course, knows what the correct response should be, especially since the stimulus will have become quite intense after a few trials and thus should be perfectly detectable, will generate response sequences that are different from the sequences generated by the anosmic observer, who lacks this knowledge. A malingerer, for example, could simply alternate

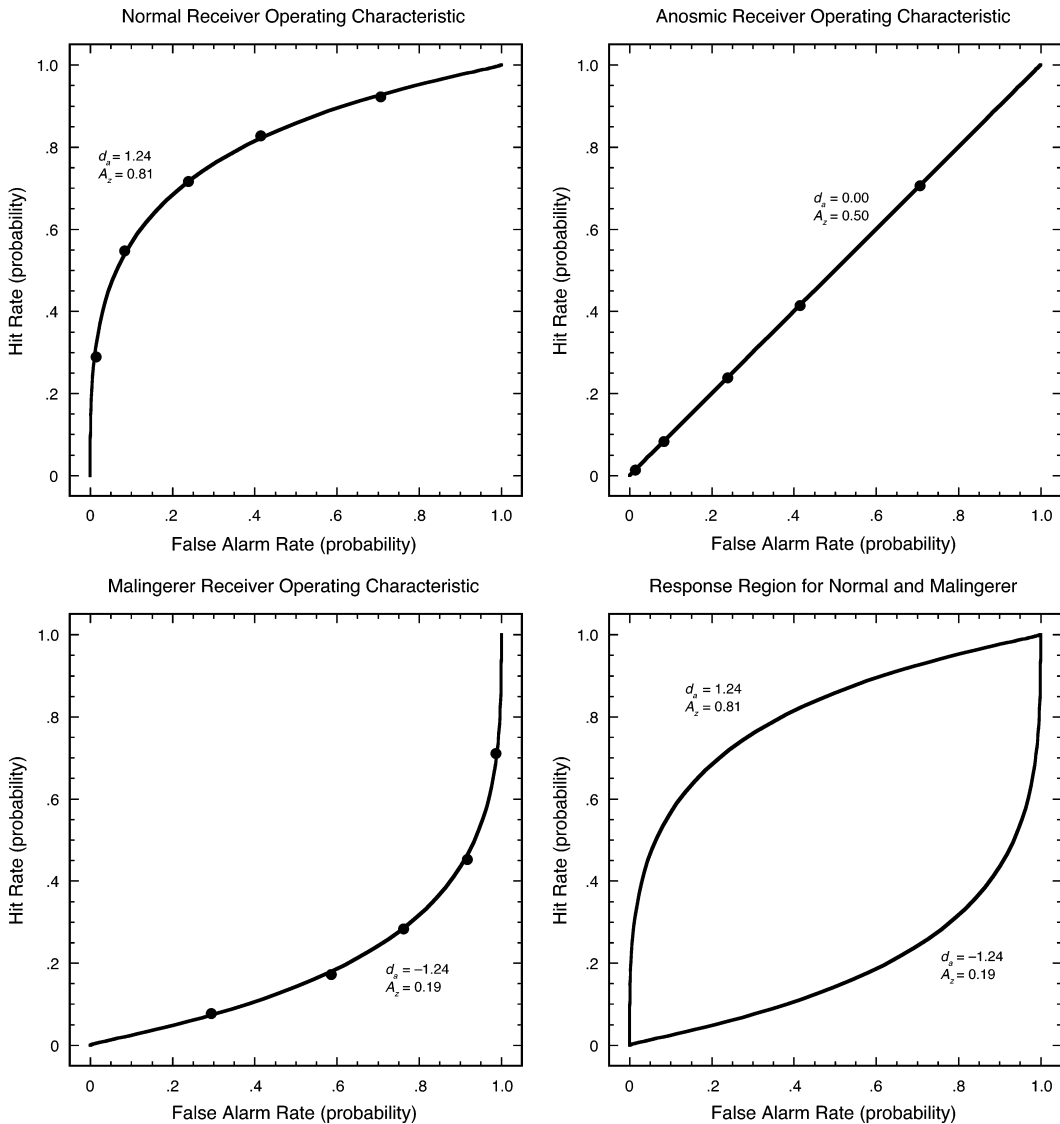


Figure 1. Receiver operating characteristics of 3 observers detecting a weak stimulus: normal (upper left panel), anosmic (upper right panel), and malingerer (lower left panel); the ROCs of the normal and of the malingerer together enclose a region of all possible hit rate-false alarm rate pairs (lower right panel).

responses (“1,” “2,” “1,” “2,” . . .) to achieve the chance detection probability of .5 in a 2AFC task that is characteristic of true anosmics. Such a sequence, however, is obviously nonrandom and would be easily detected by response-sequence measures such as adjacency, runs, and coupon. The purpose of the experiment reported here, therefore, is to combine the two traditional measures of psychophysical performance (threshold and probability of being correct) with measures of response-sequence randomness in an effort to discriminate malingerers from anosmics.

METHOD

Participants

Twenty adults participated in the experiment. They were divided into three experimental groups: (1) 7 normal participants (2 women, 5 men; mean age 31.4 years) with no known smell dysfunction, (2) 6 participants with anosmia, either of congenital or of posttraumatic origin (4 women, 2 men; mean age 41.3 years), and (3) 7 designated malingerers (3 women, 4 men; mean age 36.3 years) with no known smell dysfunction. Participants were randomly assigned to the normal and the malingering groups. The normal and malingering participants were recruited from employees at the University of Colorado Health Sciences Center. The anosmics were re-

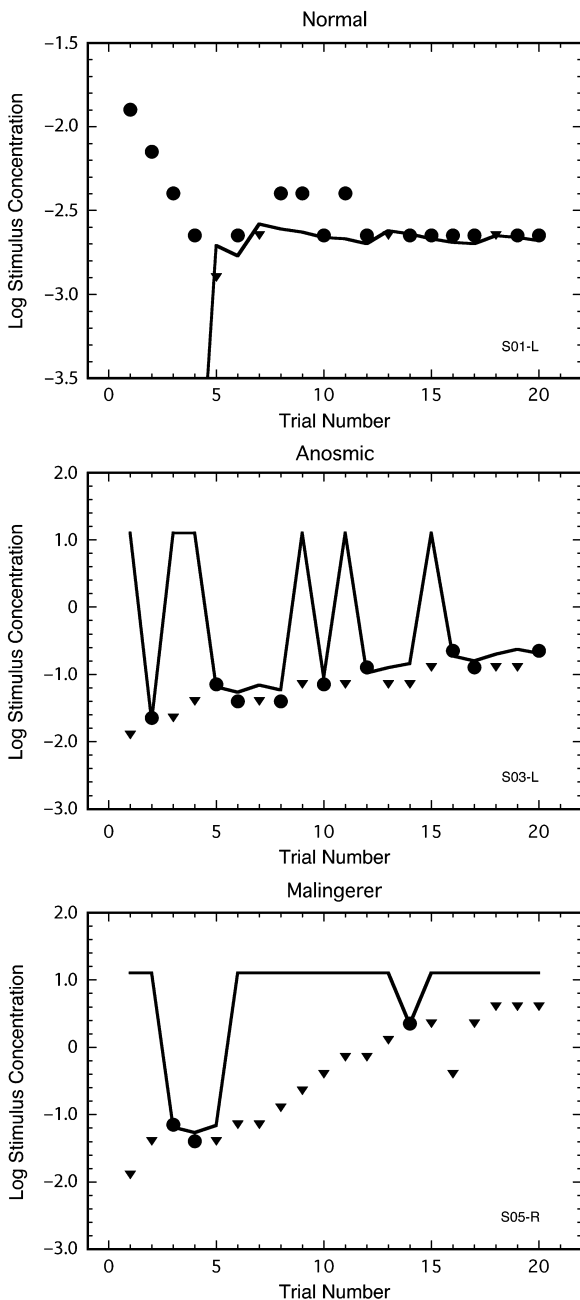


Figure 2. Behavior of maximum-likelihood adaptive-staircase method for three types of observer: normal (upper panel), anosmic (middle panel), and malingeringer (lower panel). The markers indicate the stimulus strength on each trial. Large solid circles indicate trials on which the observer was correct; small solid triangles indicate trials on which the observer was wrong. The solid line marks the maximum-likelihood estimate of the threshold concentration after each trial. Note that the vertical axis of the upper panel is different from that of the other two.

cruited from the patient population in the Denver area who had been previously examined in the clinic. All were paid to participate and were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 1992) and the ethical standards of the University of Colorado.

Stimuli

The stimuli were 20 concentrations of butyl alcohol spanning the range in quarter log steps from well above to well below the expected normal threshold of -2.5 log percent concentration. The strongest stimulus was 4% (0.60 log units), and the weakest was 0.000711% (-4.15 log units). Stimuli were presented to the participants as 60 ml in clean 250-ml amber glass bottles fitted with Teflon nosepieces (Dalton, Wysocki, Brody, & Lawley, 1997). The nosepiece consisted of a hollow Teflon rod that completely filled one nostril. Keeping the other nostril closed with a finger, the subject sniffed, drawing air from the headspace of the bottle into the nose.

Procedure

The threshold measurements were carried out single blind: The tester did not know which participants belonged to which group. Before entering the testing room, normal participants were told by the first author to go in and perform the olfactory detection task as instructed by the assistant administering the test. Anosmics were told to go in and listen to the tester's instructions and perform the task, but not comment to the tester that they were in fact unable to smell anything. Malingeringers were told to go in and listen to the tester's instructions and, by means of their performance on the psychophysical tests, to convince the tester that they were unable to smell anything. All participants were instructed to refrain from making any comments about their olfactory ability or lack thereof. They were to only answer questions and give responses. All participants complied with these instructions. With one exception, the malingeringers were tested a 2nd time to confirm that they, in fact, had normal smell sensitivity.

Thresholds were measured with a temporal 2AFC paradigm supervised by a computer program (SensoryTester) that implemented the maximum-likelihood staircase procedure. On each trial, two bottles were presented sequentially: one containing the stimulus smell and one containing distilled water. The participant sniffed the first bottle once, then the second bottle once. The participant indicated which of the two bottles had a smell different from that of water. The computer program randomly selected the interval containing the smell. If uncertain, the participant was instructed to guess. The tester entered the participant's choice into the computer. The program then computed a new estimate of the threshold and recommended the stimulus concentration for the next trial. Each nostril was tested separately in an order that alternated with participant. Detection thresholds were determined by administering 20 trials of the maximum-likelihood adaptive-staircase procedure described above (Linschoten et al., 2001). The stimulus used on the first trial was the same for all subjects: stimulus number 10 in the sequence of 20 concentrations (0.0126%; -1.90 log units). Each trial required about 40 sec. The testing of both nostrils was completed in about 30 min.

Data Analysis

The sequence of the 20 concentrations presented to the participant, the interval in which the stimulus was presented (1 or 2), the participant's response interval ("1" or "2"), and the sequence of correct and incorrect trials was analyzed by computing the following measures:

Log threshold concentration. The adaptive-staircase procedure provides a trial-by-trial running estimate of the log threshold concentration: the concentration that produces a detection probability of .75, halfway between chance performance (.50) and perfect detection (1.0). For reasons of computational stability, the highest possible threshold estimate could not be more than 0.5 log units above the strongest stimulus, and the lowest possible threshold estimate could not be lower than 0.5 log units below the weakest stimulus. This ceiling on the estimated threshold value is clearly seen in the middle and lower panels of Figure 2.

Probability correct. This measure is computed over all 20 trials by comparing the interval containing the stimulus with the response given by the participant. Since the stimulus concentration can change from trial to trial, this probability of being correct does not necessarily equal the 2AFC threshold performance of .75.

Response sequence measures. Towse and Valentine (1997) asked 94 subjects to generate random sequences of 100 numbers using the numbers 1–10. Towse and Neil (1998) then computed 45 different measures of randomness from these sequences. These measures are described in Appendix A. A principal component analysis of these 45 measures identified four principle factors: equality of response usage, short repetitions (the repeating of responses over small sequence lengths), long repetitions (the repeating of responses over longer sequence lengths), and prepotent associates (the tendency to produce stereotyped strings of responses). One additional measure of randomness, not computed by Towse and Neil, is information load (IL), a measure of the minimum amount of information needed to describe a sequence, based on mathematician Andrey Kolmogorov's concepts of complexity (Li & Vitányi, 1997) and the structural information theory of Emanuel Leeuwenberg (Buffart, Leeuwenberg, & Restle, 1981, 1983; van der Helm & Leeuwenberg, 1996). The computation of IL was carried out using the computer program PISA (van der Helm, 2000; van Lier, van der Helm, & Leeuwenberg, 1995), kindly provided by Peter van der Helm.

One set of the 46 response-sequence measures was computed from the 20-trial sequence of right or wrong, assigning the integer value of 1 to trials on which the response was correct and 0 to incorrect trials. A second set of these response-sequence measures was computed from the interval choices made by the participant on the 20 test trials (interval 1 or 2). Because our response sequences had only two values (0 and 1, or 1 and 2), and because of the small number of trials, some of the sequence measures were perfectly correlated with each other (e.g., the number correct has a correlation of -1.0 with the number incorrect), and some could not be computed at all (e.g., repetition lags greater than 10). Some measures, like coupon (mean number of trials before all responses have been given), require that each response be used at least once. One of our malingers was wrong on all 20 trials. We added a 21st correct trial to his 20 incorrect trials in order to allow the coupon index to be computed in a way that permits comparison with coupon scores of the other participants who had sequences containing both correct and incorrect trials. Including or excluding the data from this one participant did not alter the outcome of our analyses, and we therefore include him in our results.

The correlation matrices of the two sets of 46 response-sequence measures were computed, and a principal component analysis (PCA) was performed on each (SAS Institute, 1998a, 1998b) in order to identify and eliminate measures that were either perfectly correlated with each other or were so collinear that they caused a singularity during computation of the PCA. Eight PCA factors had eigenvalues greater than one, accounting for 82% of the variance in the two measurement sets. After varimax orthogonal rotation, the pattern of variable loadings on the factors was similar to that reported by Towse and Neil (1998). Preliminary discriminant analyses allowed the elimination of additional sequence measures because they did not contribute to discrimination among the groups. The correct–incorrect sequences ended up with 24 discriminant variables, and the interval sequences had 20 variables. The mean and standard deviation of each of these 24 correct–incorrect sequence measures are presented in Appendix B for the three groups of participants. The means and standard deviations for the 20 interval response sequence measures are presented in Appendix C.

In summary, each sequence of 20 trials was characterized by 46 variables: two traditional psychophysical measures (detection threshold [α] and probability of being correct [P_c]), 24 measures of correct–incorrect sequence randomness, and 20 measures of response interval sequence randomness. Discriminant analyses (van de Geer, 1971) were computed with combinations of these three sets of variables (SPSS 11.0.2, Noruésis & SPSS, 1994, 2002). The classification functions computed by the discriminant analyses reported below assumed that the three groups of participants occurred with equal a priori probability (.333).

RESULTS

Psychophysical Measures

First we present the results for the traditional psychophysical measures: log threshold concentration threshold and probability correct. The mean values for the three groups of participants are given in Table 1. Separate repeated measures analyses of variance (with nostril as the repeated measure) showed that the malingers, when not malingering, did not significantly differ from the other normal participants in the experiment. The normal, anosmic, and malingering groups, however, differed significantly ($p < .0001$) from each other. We therefore combined the malingering normal data with the data from the normal group for further analysis. The distributions of the threshold values and the probability correct for the normal, anosmic, and malingering groups are shown in Figure 3 as box-and-whisker plots. There is hardly any overlap between the values associated with normal smell function and the other two groups, but there is considerable overlap between the anosmic group and the malingering group. This overlap makes it impossible to perfectly discriminate the two groups from each other using either the threshold concentration or the probability correct variables.

This difficulty in discriminating the anosmics from the malingers was confirmed by computing a discriminant analysis with log threshold concentration (α) and probability of a correct response (P_c) as discriminant variables. The ability of the resulting two discriminant functions to classify the cases into correct groups is shown in Table 2. Although all of the normal subjects were correctly classified, only 68% percent of the malingers and anosmics were correctly classified. The position of each nostril in this 2-D discriminant-function space is plotted in the upper panel of Figure 4. The Y-shaped classification boundary in Figure 4 has its origin at the circumcenter of the triangle formed by the centroids of the three groups. From the circumcenter, each boundary line passes through the midpoint of a side of the triangle. The first discriminant dimension is able to separate the normals from both the anosmics and the malingers. The second dimension, however, is not able to separate the malingers from the anosmics very well. Overall, 84% of

Table 1
Mean Concentration Thresholds and Probability of Being Correct

Experimental Group	Number of Nostrils	Log Threshold Concentration		Probability of Correct Response	
		<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Normal	14	−2.549	.11	.732	.023
Normal malingering	11	−2.480	.11	.723	.025
Anosmic	12	−0.632	.24	.500	.033
Malingering	13	0.323	.27	.292	.052

Note—Left and right nostrils were tested separately. There were 7 normals, 6 anosmics, and 7 malingers. One malingering was tested only on the right side. One malingering did not return for a followup test to verify normosmia.

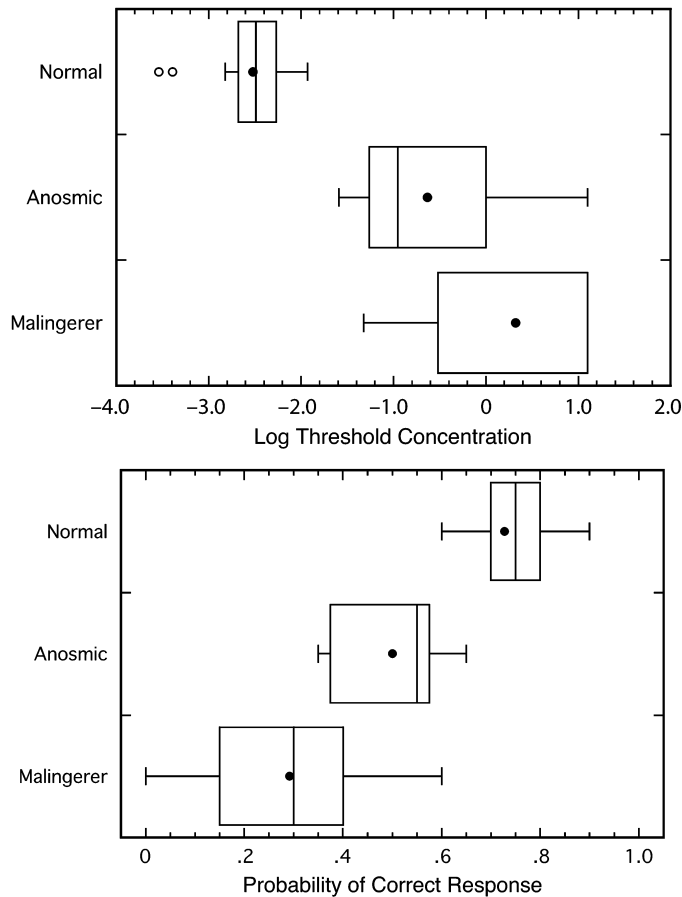


Figure 3. Box-and-whisker plots of log threshold concentration (upper panel) and the probability of a correct response (lower panel) for normals, anosmics, and malingerers. Each box encloses the data from the 25th to the 75th percentile (the interquartile distance, or IQD). The line within each box marks the median (50th percentile). The filled circle marks the mean. The end of the whiskers mark the minimum and maximum data values that lie within an “acceptable” range (1.5 IQD). Outliers are plotted with open circles.

the three groups are correctly classified, but only 68% of the malingerers and anosmics are correctly classified. All the classification errors involve anosmics incorrectly classified as malingerers (3 out of 12) and malingerers incorrectly classified as anosmics (5 out of 13). No normal subject was incorrectly classified.

Response-Sequence Measures

Two more discriminant analyses were computed by augmenting the measures α and P_c with either the 24 correct–incorrect sequence measures or the 20 response interval sequence measures. The addition of each set of the response-sequence variables improves the correct

Table 2
Classification Performance of the Two Discriminant Functions Based on Log Threshold Concentration (α) and Probability of Being Correct (P_c)

Actual Group Membership	Number of Nostrils	Predicted Group Membership					
		Normals		Anosmics		Malingerer	
		No.	% Correct	No.	% Correct	No.	% Correct
Normals	25	25	100.0	0	0.0	0	0.0
Anosmics	12	0	0.0	9	75.0	3	25.0
Malingerer	13	0	0.0	5	38.5	8	61.5

Note—Overall classification performance was 84% correct, assuming equal a priori group probability.

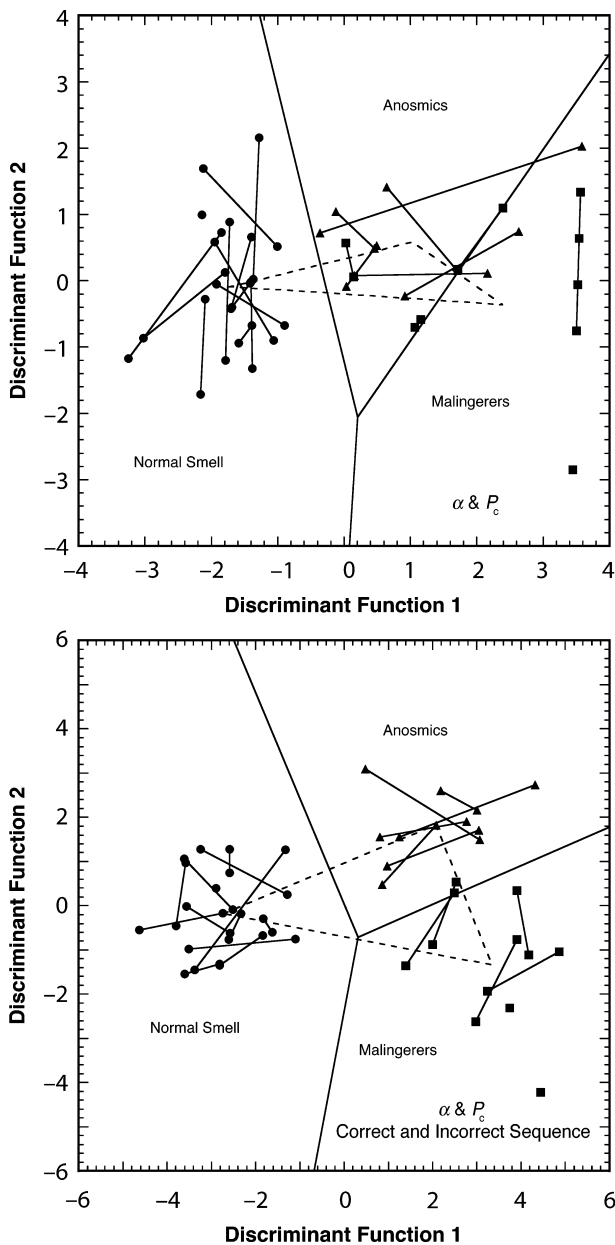


Figure 4. The position of each nostril in the 2-D discriminant analysis space for three types of participants: normals (circles), anosmics (triangles), and malingers (squares). The two symbols for the nostrils of each person are connected with a solid line. The dotted lines connect the centroids of the groups. Solid demarcation lines represent classification boundaries. The upper panel is the discriminant space based on two predictor variables (α and P_c). The lower panel is the space based on 18 predictor variables (α , P_c , and correct-incorrect response-sequence variables).

classification of malingers and anosmics from 68% to 92% and 88%, respectively. The ability of the resulting two discriminant functions to classify the cases is given in Table 3. The position of each case in the 2-D discriminant space from the discriminant analysis based on α , P_c , and the correct-incorrect sequence variables is plot-

ted in the lower panel of Figure 4. The corresponding plot based on α , P_c , and the interval sequence variables is quite similar and is not presented here. It must be noted that, although the inclusion of the sets of response-sequence variables greatly improves the discrimination of malingers from anosmics, some of the cases that were correctly classified fall very close to the boundary between the three types of participants.

A final discriminant analysis was computed with α , P_c , the 24 correct-incorrect response-sequence variables, and 11 of the 20 interval one-two response-sequence variables. The nine variables that were excluded did not contribute to discrimination: Including or removing them from the analysis had no effect on the outcome. The resulting two discriminant functions classify all cases without errors. The discriminant space for this final analysis is shown in Figure 5. The three types of cases are now closely clustered around their group centroids and fall far from the classification boundaries between groups.

The discriminant variables whose correlation with at least one of the two standardized discriminant functions is greater than 0.1 are listed in Table 4. These correlations reveal that the first discriminant function is based on threshold concentration (α) and the probability of being correct over the 20 trials (P_c). The second discriminant function is largely based on the response-sequence variables. The correct-incorrect sequence contributes somewhat more to the second discriminant function than does the interval sequence, but both are necessary to achieve perfect classification performance.

DISCUSSION

We face a dilemma publishing the techniques for detecting malingers: We risk giving malingers the critical knowledge needed to successfully avoid detection. We depend to some degree on the malingers' ignorance about the nature of psychophysical testing in order to catch them. Prior knowledge about test characteristics will influence their performance and could lead to successful malingering if only traditional psychophysical measures such as threshold value and accuracy are used.

In the present study, malingers have higher thresholds and a lower probability of being correct. The tendency to exaggerate symptoms and responses is typical of the malingering patient (Iverson & Franzen, 1998; Morel, 1998; Schmand et al., 1998). The successful malingering avoids this exaggeration (Edens et al., 2001).

The maximum-likelihood adaptive-staircase procedure has been proven to produce fast and reliable chemosensory thresholds (Linschoten et al., 2001). By using the sequence of responses given by the participant during testing, it is also helpful for detecting malingers. The strength of the approach described here lies in our use of response-sequence information because, while a malingering might be able to generate responses that give a .5 probability of being correct, it is exceedingly difficult to produce the kind of response sequences characteristic of an anosmic. This is especially true because the malin-

Table 3
Classification Performance of the Two Discriminant Functions Based on Log Threshold Concentration (α), Probability of Being Correct (P_c), and the 24 Correct-Incorrect Sequence Measures (Upper Row of Each Cell) and Performance Based on Log Threshold Concentration (α), Probability of Being Correct (P_c), and the 20 Interval Sequence Measures (Lower Row of Each Cell)

Actual Group Membership	Number of Nostrils	Predicted Group Membership					
		Normals		Anosmics		Malingers	
		No.	% Correct	No.	% Correct	No.	% Correct
Normals	25	25	100.0	0	0.0	0	0.0
		25	100.0	0	0.0	0	0.0
Anosmics	12	0	0.0	12	100.0	0	0.0
		0	0.0	11	91.7	1	8.3
Malingers	13	0	0.0	2	7.7	11	92.3
		0	0.0	2	15.4	11	84.6

Note—Overall classification performance was 96% correct for the upper row and 94% for the lower, assuming equal a priori group probability.

gerer, unlike the anosmic, knows what the correct response is on each testing trial and must try to ignore this knowledge. This knowledge leads the malingerer to have greater sequential dependencies than the anosmic—that is, the response given on trial n is more influenced by the response already given on trials $n-1$ and $n-2$. The sequential dependency is reflected, for example, in the higher RNG and RNG2 scores (see Appendix B) of the malingerer (.79 and .69, respectively) than those of the anosmic (.72 and .62, respectively). Even if the malingerer could

know these characteristics, it would be difficult or impossible to know how to modify his/her response sequence to become more like that of an anosmic. We do not know how to instruct a person to generate a sequence to imitate an anosmic.

In medico-legal cases, the threshold method described above can be fruitfully combined with suprathreshold olfactory tests such as the UPSIT or the OCM tests mentioned in the introduction in order to strengthen the case for or against malingering. Evaluations of nonsensory functioning, such as the Computerized Assessment of Response Bias (CARB) and the Word Memory Test (WMT), may also be used. Green and Iverson found that litigants with mild head injury who showed poor effort on the CARB or WMT were 4.5 times more likely to demonstrate olfactory deficits on the Alberta Smell Test than were litigants who showed adequate effort (P. Green & Iverson, 2001).

We have not provided here the specific discriminant classification function that was so successful in perfectly classifying our subjects, because we do not know how they would generalize to other populations. One should have many more than the 50 nostrils we tested to properly assess stability and reliability of the discriminant solution. What we want to provide here is motivation for carrying out more extensive investigations into the use of response-sequence statistics with a larger subject population.

It also should be noted that the 2AFC paradigm is especially useful in detecting malingerers in part because chance level performance is in the middle of the range of possible detection probabilities. The probability space where a malingerer can operate (.5–0) is just as large as the space for a normal person (1.0–.5), which makes it equally easy to detect deviations above and below chance. Contrast this symmetry with the asymmetry of a 4AFC or the UPSIT test where the probability of being correct by chance alone is .25. The operating space of a person with normal sensory capacity ranges from 1.0–.25 (a range of .75), but the operating space for a malingerer is only .25–0 (a range of .25).

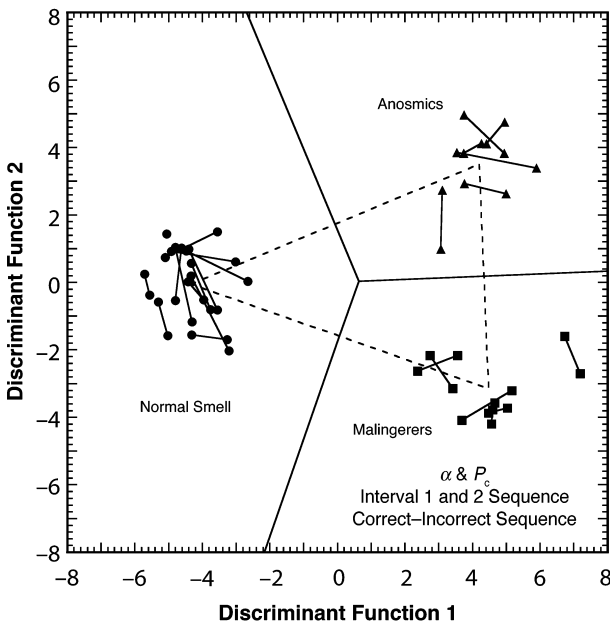


Figure 5. The position of each nostril in the final 2-D discriminant analysis space for three types of participants: normals (circles), anosmics (triangles), and malingerers (squares). The nostrils of each person are connected with a solid line. The dotted lines connect the centroids of the groups. Solid demarcation lines represent classification boundaries.

Table 4
The Correlation of Each Discriminant Variable With the Two Discriminant Functions Based on All 37 Predictor Variables, Arranged in Order of Decreasing Importance

Measure	Discriminant Function	
	1	2
Alpha	.401*	-.192
P_c	-.311*	.236
R(rw)	-.030	-.181*
RNG2(rw)	.005	-.177*
RNG(rw)	-.038	-.175*
IL(rw)	.024	.142*
RNG(12)	.042	.141*
COUPON(rw)	.025	-.139*
R(12)	-.029	.120*
NSQ(rw)	-.033	-.115*
TPI(rw)	.020	.104*
IL(12)	.015	-.103*
N_ONE(12)	-.057	-.100*

Note—Only correlations of 0.1 and higher are given. The higher of the two correlations is marked with an asterisk. Sequence measures computed from correct–incorrect sequences are indicated by (rw), and those computed from interval choice sequences are indicated by (12).

Finally, we wish to emphasize that our success in discriminating normals, anosmics, and malingers from each other is achieved with only 20 psychophysical trials per nostril. Both nostrils can be tested in about 30 min, a reasonable amount of time, given the constraints of taste and smell testing. For vision and audition, where stimuli are usually under computer control and each trial can be completed in a couple of seconds, it would be practical to administer many more than 20 trials. We anticipate that response-sequence statistics computed on longer response sequences would be able to segregate malingers from the other groups in the discriminant space even better. This method of analysis holds great promise for improving the interpretation of psychophysical results.

REFERENCES

- AMERICAN PSYCHIATRIC ASSOCIATION (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR* (4th ed.). Washington, DC: Author.
- AMERICAN PSYCHOLOGICAL ASSOCIATION (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, *47*, 1597-1611.
- ANDRIKOPOULOS, J. (2001). Malingering disorientation to time, personal information and place in mild head injured litigants. *Clinical Neuropsychologist*, *15*, 393-396.
- BAR-HILLEL, M., & WAGENAAR, W. A. (1993). The perception of randomness. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 369-393). Hillsdale, NJ: Erlbaum.
- BRUGGER, P. (1997). Variables that influence the generation of random sequences: An update. *Perceptual & Motor Skills*, *84*, 627-661.
- BUFFART, H. F. J. M., LEEUWENBERG, E. L. J., & RESTLE, F. (1981). Coding theory of visual pattern completion. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 241-274.
- BUFFART, H. F. J. M., LEEUWENBERG, E. L. J., & RESTLE, F. (1983). Analysis of ambiguity in visual pattern completion. *Journal of Experimental Psychology: Human Perception & Performance*, *9*, 980-1000.
- DALTON, P., WYSOCKI, C. J., BRODY, M. J., & LAWLEY, H. J. (1997). Perceived odor, irritation, and health symptoms following short-term exposure to acetone. *American Journal of Industrial Medicine*, *31*, 558-569.
- DELANK, K.-W., NIESCHALK, M., SCHMÄL, F., & STOLL, W. (1999). Besonderheiten in der Begutachtung von Riech- und Schmeckstörungen [Special characteristics of expert assessment of olfactory and taste disorders]. *Laryngo-Rhino-Otologie*, *78*, 365-372.
- DOTY, R. L., SHAMAN, P., KIMMELMAN, C. P., & DANN, M. S. (1984). University of Pennsylvania Smell Identification Test: A rapid quantitative olfactory function test for the clinic. *Laryngoscope*, *94*, 176-178.
- EDENS, J. F., GUY, L. S., OTTO, R. K., BUFFINGTON, J. K., TOMICIC, T. L., & POYTHRESS, N. G. (2001). Factors differentiating successful versus unsuccessful malingerers. *Journal of Personality Assessment*, *77*, 333-338.
- EVANS, F. J. (1978). Monitoring attention deployment by random number generation: An index to measure subjective randomness. *Bulletin of the Psychonomic Society*, *12*, 35-38.
- FAHLE, M., & MOHN, G. (1989). Assessment of visual function in suspected ocular malingering. *British Journal of Ophthalmology*, *73*, 651-654.
- GINSBURG, N., & KARPIUK, P. (1994). Random generation: Analysis of the responses. *Perceptual & Motor Skills*, *79*, 1059-1067.
- GREEN, D. M., & SWETS, J. A. (1974). *Signal detection theory and psychophysics* (Reprint, with corrections of the original 1966 ed.). Huntington, NY: Krieger.
- GREEN, P., & IVERSON, G. L. (2001). Effects of injury severity and cognitive exaggeration on olfactory deficits in head injury compensation claims. *NeuroRehabilitation*, *16*, 237-243.
- GUZDIOL, H., SCHUBERT, M., & HUMMEL, T. (2001). Decreased trigeminal sensitivity in anosmia. *ORL: Journal for Oto-Rhino-Laryngology & Its Related Specialties*, *63*, 72-75.
- HALL, H. V., & POIRIER, J. G. (2000). *Detecting malingering and deception: Forensic distortion analysis* (2nd ed.). Boca Raton, FL: CRC Press.
- HORNE, R. L., EVANS, F. J., & ORNE, M. T. (1982). Random number generation, psychopathology, and therapeutic change. *Archives of General Psychiatry*, *39*, 680-683.
- HUMMEL, T., BARZ, S., LÖTSCH, J., ROSCHER, S., KETTENMANN, B., & KOBAL, G. (1996). Loss of olfactory function leads to a decrease of trigeminal sensitivity. *Chemical Senses*, *21*, 75-79.
- IVERSON, G. L., & FRANZEN, M. D. (1998). Detecting malingered memory deficits with the Recognition Memory Test. *Brain Injury*, *12*, 275-282.
- KURTZ, D. B., WHITE, T. L., HORNUNG, D. E., & BELKNAP, E. (1999). What a tangled web we weave: Discriminating between malingering and anosmia. *Chemical Senses*, *24*, 697-700.
- LEOPOLD, D. A., HORNUNG, D. E., & SCHWOB, J. E. (1992). Congenital lack of olfactory ability. *Annals of Otolaryngology, Rhinology & Laryngology*, *101*, 229-236.
- LI, M., & VITÁNYI, P. (1997). *An introduction to Kolmogorov complexity and its applications*. New York: Springer-Verlag.
- LINSCHOTEN, M. R. [I.], HARVEY, L. O., JR., ELLER, P. M., & JAFEK, B. W. (2001). Fast and accurate measurement of taste and smell thresholds using a maximum-likelihood adaptive staircase procedure. *Perception & Psychophysics*, *63*, 1330-1347.
- LÖHNER, L. (1940). Über Geruchsprüfungen mit besonderer Rücksichtnahme auf echte und simulierte Anosmien [On smell testing with special emphasis on true and simulated anosmia]. *Monatsschrift für Ohrenheilkunde*, *74*, 479-495.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MOREL, K. R. (1998). Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. *Journal of Personality Assessment*, *70*, 479-495.
- NEURINGER, A. (1986). Can people behave "randomly?" The role of feedback. *Journal of Experimental Psychology: General*, *115*, 62-75.
- NICHOLSON, R. A., MOUTON, G. J., BAGBY, R. M., BUIS, T., PETERSON, S. A., & BUIGAS, R. A. (1997). Utility of MMPI-2 indicators of response distortion: Receiver operating characteristic analysis. *Psychological Assessment*, *9*, 471-479.
- NORUÉSIS, M. J., & SPSS (1994). *SPSS Advanced Statistics 6.1*. Chicago: SPSS.

- NORUÉSIS, M. J., & SPSS (2002). *SPSS 11.0 guide to data analysis*. Upper Saddle River, NJ: Prentice-Hall.
- PRIGATANO, G. P., SMASON, I., LAMB, D. G., & BORTZ, J. J. (1997). Suspected malingering and the Digit Memory Test: A replication and extension. *Archives of Clinical Neuropsychology*, *12*, 609-619.
- ROGERS, R. (Ed.) (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford.
- SAS INSTITUTE (1998a). *StatView: StatView Reference*. Cary, NC: Author.
- SAS INSTITUTE (1998b). *StatView: Using StatView*. Cary, NC: Author.
- SCHMAND, B., LINDEBOOM, J., SCHAGEN, S., HEIJT, R., KOENE, T., & HAMBURGER, H. L. (1998). Cognitive complaints in patients after whiplash injury: The impact of malingering. *Journal of Neurology, Neurosurgery, & Psychiatry*, *64*, 339-343.
- SUCHY, Y., & SWEET, J. J. (2000). Information/Orientation subtest of the Wechsler Memory Scale-Revised as an indicator of suspicion of insufficient effort. *Clinical Neuropsychologist*, *14*, 56-66.
- TOWSE, J. N. (1998). On random generation and the central executive of working memory. *British Journal of Psychology*, *89*, 77-101.
- TOWSE, J. N., & MCLACHLAN, A. (1999). An exploration of random generation among children. *British Journal of Developmental Psychology*, *17*, 363-380.
- TOWSE, J. N., & NEIL, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments, & Computers*, *30*, 583-591.
- TOWSE, J. N., & VALENTINE, J. D. (1997). Random generation of numbers: A search for underlying processes. *European Journal of Cognitive Psychology*, *9*, 381-400.
- TULVING, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, *69*, 344-354.
- TUNE, G. S. (1964). A brief survey of variables that influence random-generation. *Perceptual & Motor Skills*, *18*, 705-710.
- VAN DE GEER, J. P. (1971). *Introduction to multivariate analysis for the social sciences*. San Francisco: Freeman.
- VAN DER HELM, P. A. (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, *126*, 770-800.
- VAN DER HELM, P. A., & LEEUWENBERG, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review*, *103*, 429-456.
- VAN DER HELM, P. A., VAN LIER, R. J., & LEEUWENBERG, E. L. J. (1992). Serial pattern complexity: Irregularity and hierarchy. *Perception*, *21*, 517-544.
- VAN LIER, R. J., VAN DER HELM, P. A., & LEEUWENBERG, E. L. J. (1995). Competing global and local completions in visual occlusion. *Journal of Experimental Psychology: Human Perception & Performance*, *21*, 571-583.
- WAGENAAR, W. A. (1968). Sequential response bias in psychophysical experiments. *Perception & Psychophysics*, *3*(5B), 364-366.
- WAGENAAR, W. A. (1970). Subjective randomness and the capacity to generate information. *Acta Psychologica*, *33*, 233-242.
- WAGENAAR, W. A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin*, *77*, 65-72.
- WICKENS, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.

APPENDIX A

The response-sequence measures used in the discriminant analyses are described in detail by Towse and Neil (1998), with the exception of Information Load (IL) (Li & Vitányi, 1997; van der Helm, 2000; van Lier et al., 1995).

Redundancy (R). A measure of the probability of each response relative to that expected by chance. A score of 0% indicates each response occurred equally often; A score of 100% indicates that only one response was used.

Random number generation (RNG). A measure of the dependency of a response on the previous one. A score of 0 indicates no dependency; A score of 1 indicates complete dependency.

Random number generation 2 (RNG2). A measure of the dependency of a response on the second previous one. A score of 0 indicates no dependency; A score of 1 indicates complete dependency.

Null-score quotient (NSQ). The total number of response pair permutations that do not appear in the response sequence relative to the total possible pairs. A score of 0 means that all possible pairs were used; A score of 100 means that no pairs were used.

Turning point index (TPI). The total number of trials that mark transitions between ascending and descending sequences of responses divided by the number expected in a truly random sequence. Values greater than 100% indicate too many turning points; values less than 100% indicate too few.

Runs (RUNS). The variance in the number of items in successive ascending sequences.

Coupon (CP). Mean number of responses produced before all possible responses have been given.

Adjacency (A). The percentage of the total number of response pairs in the response sequence that are adjacent pairs (e.g., 0,1 and 1,0).

First-order differences (FOD-0). The number of sequential response pairs whose responses are the same. Other first-order differences (i.e., -1 and +1) are perfectly correlated with the Adjacency measure and are therefore not included here.

Repetition distance frequencies (REP_1 through REP_20). The frequencies of each possible number of trials required to repeat a response. REP_11 through REP_20 were zero in the data set and were not used.

Mean repetition gap (RG). The mean number of trials needed for each possible response to repeat itself.

Median repetition gap (RG). The median number of trials needed for each possible response to repeat itself.

Modal repetition gap (RG). The modal number of trials needed for each possible response to repeat itself.

Phi index (PHI_2 through PHI_7). A measure of the tendency to repeat a response over response sequences of different lengths. It is derived from the chi-square statistic comparing the observed frequencies with the expected frequencies. A positive value indicates more response repetition than expected; a negative value indicates less repetition than expected.

Information load (IL). A measure of how much information is needed to completely describe a response sequence. Random sequences require the maximum amount of information, whereas a repeating sequence of alternations (e.g., 1, 2, 1, 2, . . .) require the minimum description.

APPENDIX B
Group Means and Standard Deviations of the
24 Correct–Incorrect Response–Sequence Variables
Used in the Discriminant Analysis

Variable	Normals		Anosmics		Malingers	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
R	18.1	13.0	3.53	2.8x1	22.8	25.9
RNG	0.77	0.047	0.72	0.02	0.79	0.083
NSQ	9.33	15.28	0.00	0.00	10.3	16.0
RNG2	0.66	0.046	0.62	0.019	0.69	0.09
TPI	57.7	17.3	69.4	17.2	54.5	31.5
RUNS	0.11	0.10	0.098	.12	0.096	0.099
COUPON	3.99	0.97	3.02	0.82	6.24	6.65
A	40.4	10.6	48.3	11.2	39.9	18.4
REP_2	4.00	1.63	4.08	2.15	3.38	2.69
REP_3	1.36	1.29	2.58	1.73	1.62	1.71
REP_4	0.52	0.77	0.83	0.72	0.69	0.63
REP_5	0.40	0.58	0.25	0.45	0.62	0.65
REP_6	0.28	0.54	0.33	0.49	0.077	0.28
REP_7	0.08	0.27	0.08	0.29	0.00	0.00
REP_8	0.08	0.27	0.00	0.00	0.00	0.00
REP_9	0.12	0.33	0.08	0.29	0.00	0.00
MED_GAP	1.22	0.29	1.38	0.43	1.31	0.43
MODE_GAP	1.00	0.00	1.16	0.39	1.15	0.38
PHI_3	-3.21	14.53	-8.89	17.6	-8.84	13.31
PHI_4	-5.94	11.80	-6.47	13.71	-3.83	12.01
PHI_5	-7.37	10.28	-3.95	6.29	-6.35	10.88
PHI_6	-3.76	7.37	-5.13	8.58	-0.87	7.49
PHI_7	-2.90	6.59	-4.98	8.25	-0.79	5.49
IL	6.44	1.23	7.58	1.16	6.08	2.43

APPENDIX C
Group Means and Standard Deviations of the 20 Interval
Response–Sequence Variables Used in the Discriminant Analysis

Variable	Normals		Anosmics		Malingers	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
R*	3.83	4.35	4.66	6.72	.95	1.73
RNG*	.72	.018	.74	.03	.72	.02
NSQ*	0.00	0.00	2.78	9.62	0.00	0.00
RNG2*	.62	.034	.62	.03	.60	.01
TPI*	70.3	14.0	71.5	24.5	72.4	21.6
RUNS	.12	.13	.11	.11	0.099	.12
COUPON	2.93	.43	3.14	.84	3.06	.81
A*	49.6	7.90	50.0	16.0	50.8	13.8
N_ONES*	10.48	2.28	8.75	2.26	10.1	1.19
REP_2*	4.28	2.28	4.42	2.75	4.77	2.42
REP_3	2.28	1.28	2.25	1.71	1.77	1.83
REP_4	1.12	1.23	1.00	1.04	1.31	1.18
REP_5*	.44	.65	.42	.67	.31	.63
REP_6	.16	.37	.25	.45	.31	.48
REP_7*	.12	.33	0.00	0.00	.077	.28
REP_8	.04	.20	.083	.29	.15	.38
REP_9	0.00	0.00	.083	.29	0.00	0.00
MODE_GAP	1.04	.54	1.25	.45	1.15	.38
PHI_5	-7.76	9.90	-3.57	6.46	-5.06	9.41
IL*	7.4	.91	7.17	.94	7.85	1.14

*Variable used in the final discriminant analysis.