

Normative or Confirmative: Effects of Information Quality and Redundancy in Decision Support Systems

Human Factors
2025, Vol. 67(6) 546–559
© 2024 Human Factors
and Ergonomics Society



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00187208241302787
journals.sagepub.com/home/hfs



Yoav Ben Yaakov¹ , Maja Denisova² , Filmona Mulugeta², and Joachim Meyer¹

Abstract

Objective: The study investigates users' tendency to access decision support (DS) systems as a function of the correlation between the DS information and the information users already have, the ongoing interaction with such systems, and the effect of correlated information on subjective trust.

Background: Previous research has shown inconclusive findings regarding whether people prefer information that correlates with information they already have. Some studies conclude that individuals recognize the value of noncorrelated information, given its unique content, while others suggest that users favor correlated information as it aligns with existing evidence. The impact of the level of correlation on performance, subjective trust, and the decision to use DS remains unclear.

Method: In an experiment ($N = 481$), participants made classification decisions based on available information. They could also purchase additional DS with different degrees of correlation with the available information.

Results: Participants tended to purchase information more often when the DS was not correlated with the available information. Correlated information reduced performance, and the effect of correlation on subjective trust and performance depended on DS sensitivity.

Conclusion: Additional information may not improve performance when it is correlated with available information (i.e., it is redundant). Hence, the benefits of additional information and DS depend on the information the system and the operator use.

Application: It is essential to analyze the correlations between information sources and design the available information to allow optimal task performance and possibly minimize redundancy (e.g., by locating sensors in different positions to capture independent data).

Keywords

correlated information, human–computer interaction, decision making

¹Tel Aviv University, Israel

²Berlin Institute of Technology, Germany

Received: February 6, 2023; accepted: October 31, 2024

Corresponding Author:

Joachim Meyer, Industrial Engineering, Tel Aviv University, Wolfson Building, Ramat Aviv, Tel Aviv 69978, Israel;
e-mail: jmeyer@tau.ac.il

Introduction

Decision Support (DS) systems assist operators in numerous tasks. For instance, clinical DS helps to detect brain tumors (Gupta et al., 2018), and financial DS helps investors and traders decide to buy or sell stocks (Kraus & Feuerriegel, 2017). Even though such systems should improve performance, users tend to make nonoptimal use of DS, which limits their benefits (Bartlett & McCarley, 2017; Ben Yaakov et al., 2019; Meyer, 2001; Parasuraman & Riley, 1997). Users must assign optimal weights to the information the DS provides to derive the maximal benefits from an aid. With nonoptimal weights (i.e., weights that do not accurately reflect the reliability or value of the DS information), performance may drop dramatically (Meyer & Kuchar, 2021). However, people tend to give excessive weight to nonvalid aids (over-rely) and insufficient weight to valid aids (under-rely) (Lee & See, 2004). In the worst case, operators may perform a task more poorly when assisted by a decision aid than when unassisted (Alberdi et al., 2004).

The benefits of the information from the DS depend not only on its reliability and the user's trust in it (Parasuraman et al., 2008; Parasuraman & Wickens, 2008; Schaefer et al., 2016) but also on the correlation between the information the DS provides and the user's existing information. The DS might offer redundant information, reducing its added value. In contrast, if the DS is based on independent or uncorrelated observations, the combination of both sources provides better information (Elvers & Elrif, 1997).

For instance, correlated observations arise in medical diagnostics when a physician and a DS rely on the same measurements, possibly leading to similar conclusions and making the concurrent use of both redundant. Conversely, with uncorrelated observations, the DS and the physician consider different information (e.g., genetic markers and observed symptoms). Decisions based on two uncorrelated information sources are likely better than those based on two evaluations of the same source.

The correlation between human and automated diagnoses can be interpreted in two ways. In the first case, which is the focus of this study, both rely on the same predictive information, leading to an agreement that does not indicate good predictive

value and accurate diagnoses. In the second case, independent sources consistently yield similar results, suggesting that the two sources are valid and should be trusted and integrated.

Two questions arise regarding the interaction with correlated DS: (1) How does correlated information affect task performance and subjective trust in DS? (2) Do users prefer assistance from a system that provides correlated or uncorrelated information besides the information they already have? If sources are correlated, what level of correlation do they prefer?

Regarding the second question, normatively, two (or more) nonredundant, uncorrelated sources of information should be preferred to minimize forecast error and to prevent shared biases (Bonaccio & Dalal, 2006; Hogarth, 1989; Yaniv, 2004), as each contains unique information. However, when two credible sources disagree, the conflict might complicate or delay the decision process. In previous studies, people correctly recognized the value of nonredundant information (Ben Yaakov & Meyer, 2024; Goethals & Nelson, 1973; Gonzalez, 1994). Still, people may only realize the redundancy if it is apparent in the prediction situation (Maines, 1990). Other studies suggest that decision makers prefer redundant cues and decide more confidently with them (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974). In one study (Kahneman & Tversky, 1973), participants predicted grade point averages from either a pair of highly correlated aptitude tests or uncorrelated tests. Although higher predictive accuracy could be achieved with the uncorrelated pair of tests, participants were more confident in their predictions when using the correlated tests, irrespective of the actual predictive accuracy. A later study reiterated the idea that people trust predictions based on redundant input variables more, even though such predictions are often less accurate (Tversky & Kahneman, 1974). Soll (1999) noted that it is unclear whether confidence in the highly correlated cues resulted from a preference for redundancy (multiple overlapping sources that provide similar information), consistency (reliability of cues in providing the same information under similar conditions), or a combination of both.

Automation can help to improve performance (Bartlett & McCarley, 2017; Ben Yaakov et al., 2020, 2021, 2024; Dixon et al., 2007; Parasuraman

& Riley, 1997), and the improvement increases the more reliable the system is (Botzer et al., 2013; Rieger & Manzey, 2022). The effect of correlated information on performance is unclear. Empirical studies present conflicting findings. While some claim that with correlated information, performance is worse than with uncorrelated information (Elvers & Elrif, 1997), others argue that correlated information does not affect performance (Munoz Gomez Andrade et al., 2022).

The redundancy of the information from a DS can also impact the perceived credibility of information. Hogarth (1989) showed that low-redundancy sources may seem more credible than high-redundancy sources.

While DS sensitivity and redundancy have been studied independently, the interaction between them has not been explored previously. The effect of redundancy on the tendency to access DS, the performance when using it, and the trust in it may vary depending on the sensitivity of the DS. Thus, DS sensitivity may modulate the impact of redundant information. This might explain the inconclusive findings regarding the effect of correlated information on the tendency to access DS and the performance when using it.

In a previous study, we investigated the effect of correlated information on the tendency to purchase DS, its use, and trust in it (Mulugeta et al., 2022). Participants performed a classification task for which they could purchase advice from DS, which provided information that was either noncorrelated, medium correlated, or fully correlated to the information they already had. Participants preferred to purchase the noncorrelated DS, performed better with it, and trusted it more than the medium or fully redundant systems. However, the results were inconclusive regarding the effect of partial redundancy on performance and subjective trust.

In our current study, we aim to address this limitation by examining a wider range of intermediate correlation levels (low and high correlation levels), which can shed light on the effect of intermediate levels of correlation on task performance and subjective trust. We also manipulated DS sensitivity with three levels (low, medium, and high), whereas our previous research only considered a medium level. This additional manipulation enables us to explore the interaction between DS sensitivity and redundancy, allowing us to study

their joint effects on decision outcomes and trust in DS. We expect that redundancy will have a stronger negative effect on performance as DS sensitivity increases. Similarly, redundancy with a high-sensitivity DS may lower subjective trust, as users may see it as adding little value, with a low-sensitivity DS, more redundancy may raise subjective trust.

Modeling Optimal Decisions

Many researchers of binary DS have adopted Robinson and Sorkin's (1985) contingent criterion (CC) model (Bartlett & McCarley, 2017; Maltz & Meyer, 2001). This model, rooted in signal detection theory (SDT) (Green & Swets, 1966), assumes decision makers (human or system) encode evidence for or against two mutually exclusive states, signal present and signal absent (i.e., noise), with overlapping evidence. The signal noise to be detected could represent faulty or intact products in a quality control task, the necessity of medical intervention, etc.

The performance in such tasks depends on two variables: the decision maker's sensitivity (d'), that is, the ability to distinguish between signal and noise, and the response criterion (C). The decision maker classifies events by comparing a continuous observed value with the response criterion. Values below the criterion indicate a judgment of noise, while values above it indicate signal. The decision maker's criterion can be conservative (biased toward noise with $C > 0$), liberal (biased toward signal with $C < 0$), or unbiased. A value of $d' = 0$ indicates no ability to distinguish between signal and noise. As d' increases, the ability to differentiate between the two entities increases (see Macmillan & Creelman, 2004, for calculation of d' and C).

In the CC model, the DS and the human operate sequentially. Based on its observed value, the DS generates an output, which can, for instance, be an alarm when the judgment is signal and no alarm when the judgment is noise. This output determines the operator's response criterion, with the operator adopting a more liberal criterion after a signal output from the DS and a more conservative criterion after a noise output.

The sequential process involving both DS and a human can enhance the ability to differentiate

between signal and noise events. The joint sensitivity of the DS and the human is influenced by the individual sensitivities and the correlation between the values each observes. If the human and DS observed values stem from the same source (fully correlated), the combined detection should not be better than the better detector acting alone. However, if the sources are uncorrelated, each contributes unique information, improving overall detection sensitivity.

The combined sensitivity (also called effective sensitivity) is the equivalent of a single detector with the same performance level as the human and the DS performing jointly. By definition, the effective sensitivity ($d'_{\text{effective}}$) is greater than or equal to d'_{Human} and d'_{DS} , and the maximum effective sensitivity is $d'_{\text{max}} = \sqrt{d'^2_{\text{human}} + d'^2_{\text{DS}}}$ (see Pollack & Madans, 1964) for two detectors that observe continuous stimuli and is $d'_{\text{max}} = \sqrt{d'^2_{\text{human}} + d'^2_{\text{DS}} - .3d'_{\text{human}}d'_{\text{DS}}}$ for two detectors, when DS provides a binary indication (Meyer & Kuchar, 2021). The maximum effective sensitivity decreases with higher levels of correlation between the observed values of the two detectors.

Experiment

The experiment investigated users' tendency to use DS with different redundancy and sensitivity levels and the ongoing interaction with a DS as a function of the redundancy level and sensitivity of the DS. We also aim to study the relation between the decision to use DS, task performance, and subjective trust in it.

Participants were asked to classify binary events that simulated a physician's decision on whether a patient needed medical intervention (signal) or not (noise). Participants saw a real number with two decimal places (participants' observed value), representing the seriousness of the patient's medical condition and serving as the basis for their decision. At the beginning of certain experimental blocks, participants could purchase a Clinical Decision Support System (the DS) to get additional information about the patient's medical condition for all trials in the block. This information was provided as a binary indication (DS judgment), indicating whether the patient required medical intervention (signal) or not (noise).

Participants

We recruited 481 participants from the participant sourcing platform Cloud Research©. All participants received a base payment of US\$ 1.5 for their participation. To include only participants who paid attention to the task, we excluded 25 participants with $d' = 0$ or lower in one of the analyzed, resulting in a final sample of 456 participants.

The research complied with the American Psychological Association Code of Ethics and was approved by the Tel Aviv University Ethics Committee for Research with Human Participants. Informed consent was obtained from each participant.

Apparatus

We developed a web-based experiment platform consisting of a back-end implemented with a Python infrastructure (using the Django library), a database (MySQL), and a front-end developed in HTML. Participants performed the experiment on their own devices, which were restricted to a desktop or laptop computer.

Design and Procedure

We manipulated the DS sensitivity and the correlation between the DS and the participants' observed values in a 3×5 between-subjects factorial design. Each participant classified 180 events across 6 blocks. Participants were randomly assigned to one of 15 between-subjects experimental conditions, resulting from 3 levels of DS sensitivity (d') low ($d' = 1$, *accuracy* = .68), medium ($d' = 1.5$, *accuracy* = .77), or high ($d' = 2$, *accuracy* = .84), and 5 levels of redundancy (i.e., correlation between the participants' observed value, and DS's observed value): $r = 0$ (no correlation), $r = .25$ (small correlation), $r = .5$ (medium correlation), $r = .75$ (high correlation), and $r = 1$ (full correlation).

Participants observed value was sampled from one of two normal distributions with means of 5.25 for signal absent (noise; no intervention needed) and 6.75 for signal present (signal; intervention needed). Both distributions had a standard deviation of 1, so the best-achievable d' , as determined by the variability in the participant's observed value, was 1.5. It was displayed on a [2,10] scale with two decimal places. Participants

were informed that a higher value represents a higher probability of the patient needing a medical intervention.

We manipulated the correlation between the DS's and the participants' observed values with the equation: $Y = r \cdot X + \sqrt{1 - r^2} \cdot z$ (Kaiser & Dickman, 1962). In it, r is the desired correlation, X is the participants' observed value, z is a random variable independent of X , and Y is the correlated DS stimulus (DS's observed value). This manipulation adjusts the correlation through the extent to which the DS and the participants' observed values rely on the same underlying information source, ranging from $r = 0$ (no correlation) to $r = 1$ (full correlation). When $r = 1$, $y = x$ (the DS's observed value equals the participants' observed value), and when $r = 0$, $y = z$ (the DS value is random). Z values for noise (no intervention) events were sampled from a Gaussian distribution with a mean of 5.5, 5.25, and 5, for the low, medium, and high DS sensitivity, respectively. Z values for signal (need intervention) events were sampled from Gaussian distributions with means of 6.5, 6.75, and 7 for the low, medium, and high DS sensitivity, respectively. All distributions had a standard deviation of 1. Thus, the d' was 1 for low DS sensitivity, 1.5 for medium sensitivity, and 2 for high sensitivity. The DS's observed value (Y) was converted from a continuous variable to the binary judgment using an unbiased decision threshold so that for values of 6 and above, the DS judgment was signal, and a red rectangle was shown (an intervention was needed), and for values below 6, the DS judgment was noise, and a green rectangle was shown (no intervention was needed).

Figure 1 represents typical classification screens. The dependent variables were the decision to purchase the DS (denoted as purchase decision, pd), task performance (effective sensitivity and cumulative score), and subjective trust.

After each classification, participants were informed whether their response was correct or incorrect, and their score was updated according to the scoring method explained in the instruction screens. Participants purchased aid access with a point system. They began each block with a budget of 30 points. They earned 3 points for each trial they classified correctly. They lost 6 points if they missed the need for an intervention and 3 points when they intervened when it was unnecessary. They were charged 5 points for each block in which they chose to purchase the DS. Participants were incentivized to gain maximum points as they were told they could earn additional money based on their score.

The values shown to participants and DS judgments (if accessed) were pregenerated and were similar for all participants (participants' observed value) and all participants within the same condition (DS judgment). The order of the trials within each block was randomized between subjects. The probability that a patient needed a medical intervention (signal event) was $p = .3$ (12 signal events and 28 noise events in a block).

The experiment consisted of an instruction section and six experimental blocks. The first two practice blocks (10 trials each) familiarized participants with the task and the DS, allowing them



Figure 1. Typical screens (from left to right) without DS, with DS and no intervention needed, and with DS and intervention needed. Participants observed value, DS judgment (if purchased), current score, total number of trials in the current block, and number of trials classified in the current block can be seen in it.

to form an impression of its redundancy and sensitivity. The first block did not include the DS; in the second block, the DS judgment was displayed by default. Blocks 3 to 6 were the actual experiment. Participants could choose at the beginning of each block whether to purchase information from the DS or not (if purchased, the information was received for the entire block). They then classified 40 events in each block based on the observed value and the DS judgment, if purchased. Half of the participants could not purchase the DS in block 4, and the other half could not purchase it in block 5. These blocks provided a base level of unaided task performance. We only analyzed the three blocks in which the purchase option existed. We will refer to these blocks as test blocks A, B, and C. The experiment flow is depicted in Figure 2.

At the end of each block in which they purchased the DS, participants indicated their subjective trust in the support on a scale from 1 to 10 (by answering “How much did the system help you in the task?”) and their subjective assessment of the system performance with the same scale (by answering “How good was the system in distinguishing between cases in which an intervention was needed or not needed?”).

Figure 3 depicts the different correlation levels between the human and the DS’s observed values and judgment (i.e., redundancy levels). The upper panel depicts the DS observed value, and the lower panel depicts the DS judgment, which was converted with unbiased decision thresholds. The visualizations highlight how the correlation between the human and the DS’ observed value affects the total information. Lower correlations result in more unique information and, thus, lower redundancy, while higher correlations result in less

unique information and higher redundancy. It demonstrates the trade-off between uniqueness and consistency in decision making. Lower redundancy levels broaden the range of perspectives and potential insights, while higher redundancy levels reinforce consistent information.

As the correlation between the two information sources increases, the overlap in the information they provide also increases, resulting in redundancy, which effectively reduces the total informational gain from both sources. When two information sources are completely redundant ($r = 1$), the mutual information is maximal so that a source provides no additional information beyond what is already given by the other. On the other hand, with no correlation at all ($r = 0$), the mutual information is minimal, meaning each source provides unique information that reduces uncertainty maximally. Given the correlation coefficient r , the mutual information $I(X;Y)$ for jointly Gaussian variables is: $I(X;Y) = -\frac{1}{2} \log(1 - r^2)$ (Cover & Thomas, 2006). For the specific correlation levels (0, 0.25, 0.5, 0.75, 1) in Figure 3, the corresponding mutual information values are (0, .01, .06, .18, $\rightarrow \infty$).

The DS sensitivity depicted in Figure 3 ($d' = 1$), was the same across all redundancy levels. The sensitivity reflects the number of correct and wrong classifications. Figure 3 shows that at high redundancy conditions, False Positive (+) and False Negative (°) errors aligned with participants’ observed values—False Positives occurred with high observed values, and False Negatives with low observed values. In contrast, these errors were not correlated with observed values at low redundancy levels. Consequently, detecting DS mistakes in high redundancy levels was more difficult because they aligned with the participants’ observed value.

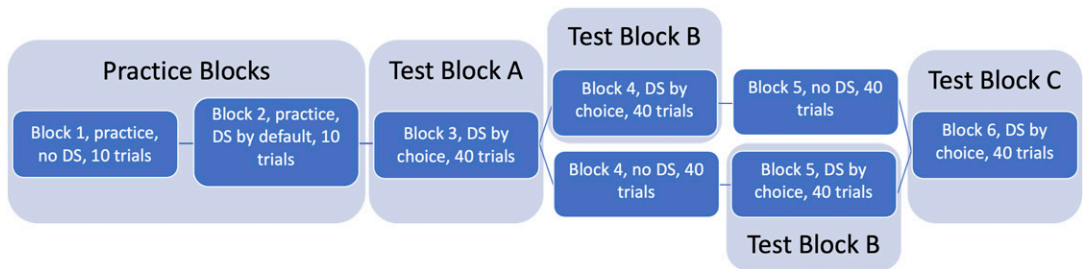


Figure 2. Experiment flow.

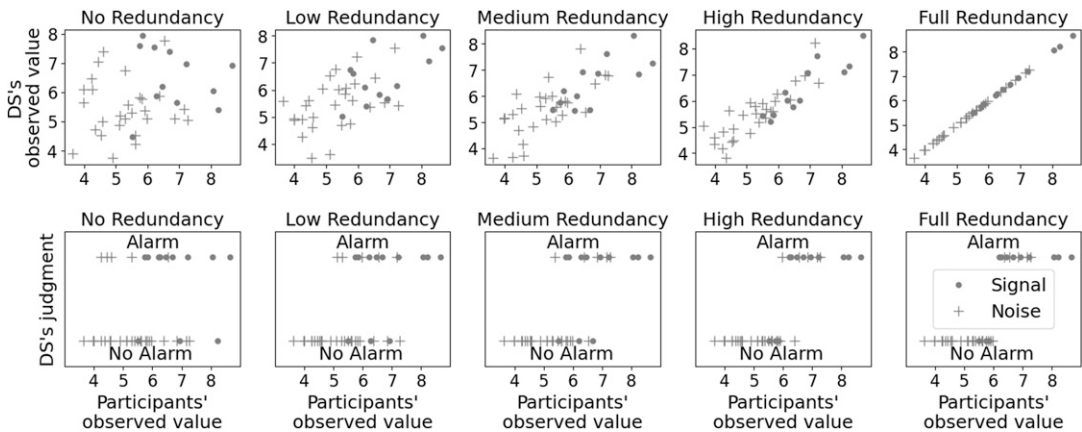


Figure 3. Participants' observed value versus DS's observed value (upper panel) and DS' judgment (lower panel) for Signal (o) and Noise (+) events.

Results

Purchase Decisions

We analyzed the purchase decisions with a Mixed Effects Logistic Regression with system sensitivity (d') and redundancy level as fixed effects, experimental block as a repeated measures variable, and participants' ID as a random effect. The interaction between system d' and redundancy level was not significant ($p = .16$). To maintain a simpler model, we excluded the nonsignificant interaction. The odds for participants to purchase the system decreased with higher redundancy, $b = -1.5$, $p < .001$, $SE = .17$, Odds Ratio (OR) = 0.22, 95% CI [.16 – .31]. Conversely, the odds of purchasing the system increased with a higher system d' , $b = 1.61$, $p < .001$, $SE = .15$ OR = 5, 95% CI [3.69 – 6.78]. Block was not a significant predictor for purchase decision (PD). The interaction between the system d' and redundancy was not significant, so the participants' tendency to purchase low redundancy DS was similar for all sensitivity levels. Figure 4 shows the main effects of system d' and redundancy level on PD.

To ensure that counterbalancing the option to purchase the DS between Blocks 4 and 5 did not influence participants' behavior, we conducted a z-test for two proportions using Python's statsmodels library (Seabold & Perktold, 2010). This test compared the purchase rates between participants who were given the option in Block 4 and Block 5. 123 of 220 participants (.56) chose to

purchase the DS in Block 4, and 142 of 236 participants (.6) did so in Block 5. There was no significant difference in the proportions of participants who purchased the DS between the two blocks ($z = -0.92$, $p = 0.29$).

Performance

We analyzed performance with both effective sensitivity (d'_{eff}) and score because they capture different aspects of participant performance. Effective sensitivity measures their accuracy in discriminating between signal and noise and the score reflects the ability to maximize rewards. Although the variables are correlated ($r = 0.96$, $p < .001$, $N = 1367$), their inclusion allows for an evaluation of different dimensions of participants' performance. We used a Generalized Linear Mixed Model (GLMM) to analyze the performance in each block with participant ID as a random effect and DS redundancy, system sensitivity (d'), and purchase decision (pd) as fixed effects. The pd, which was a dependent variable in the previous analysis, is a key factor here, and we focused on its interactions with the other fixed effects. Table 1 and Figure 5 summarize the results of the models.

Effective Sensitivity. The effective sensitivity was significantly higher when DS was purchased ($M = 1.83$; $SD = 0.34$; $N = 821$) than when it was not purchased ($M = 1.53$; $SD = 0.19$; $N = 547$). Moreover, there was a learning effect, as the effective

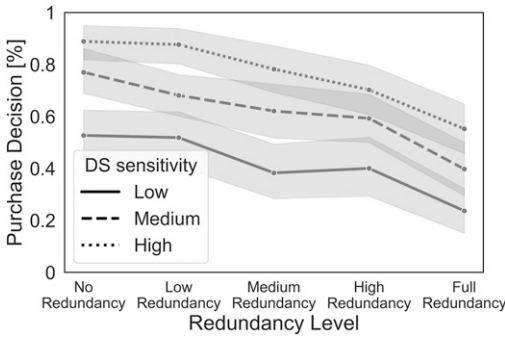


Figure 4. Mean purchases of DS by system d' and redundancy. The error ranges represent standard errors.

sensitivity increased as participants gained experience in the task ($M_1 = 1.66, SD_1 = 0.34; M_2 = 1.71, SD_2 = 0.31; M_3 = 1.76, SD_3 = 0.33$).

For a further analysis regarding the effect of pd and the significant interactions, we conducted separate analyses for blocks with and without DS. In blocks where participants decided not to purchase DS, only the effect of the block was significant, $b = .05, p < .001, SE = .01, 95\% CI [.03 - .07]$. A post-hoc analysis using the Tukey HSD test revealed a significant difference between test block A ($M = 1.47; SD = 0.19$) and test block C ($M = 1.58; SD = 0.2$), $p < .001$. System d' ($p = .16$), redundancy level ($p = .29$), and the interaction

Table 1. GLMM Model With Effective Sensitivity and Cumulative Score as Dependent Variables, Participant's ID as a Random Effect, Purchase Decision (pd), Redundancy, and System d' as Fixed Effects, and the Block as a Repeated Measures Variable.

	Effective Sensitivity			Score		
	b	SE b	95% CI	b	SE b	95% CI
Baseline	1.46***	.04	(1.39, 1.54)	87.77***	1.27	(85.28, 90.25)
Pd	-.62***	.04	(-.73, -.5)	-23.69***	2.02	(-27.63, -19.74)
System d'	-.01	.02	(-.06, -.03)	-.82	.79	(-2.38, .73)
Redundancy	.00	.025	(-.05, .05)	.59	.87	(-1.11, 2.29)
Block	.04***	.01	(.03, .06)	1.54***	.22	(1.1, 1.97)
System d' X pd	.72***	.04	(.64, .79)	22.62***	1.29	(20.1, 25.14)
Redundancy X pd	.24*	.09	(.05, .42)	2.08	3.17	(-4.13, 8.28)
System s' X redundancy X pd	-.48***	.05	(-.59, -.37)	-11.4***	1.85	(-15.03, -7.77)

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. Interactions with pd were also included in the models. b represents the estimated coefficient of the predictor.

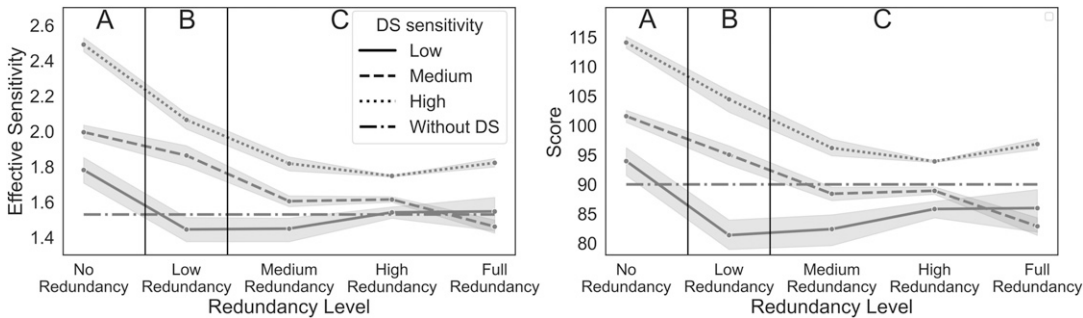


Figure 5. Average effective sensitivity (left) and cumulative score (right) dependent on system d' , redundancy level, and PD. The error ranges represent standard errors. To streamline the reporting of differences between redundancy levels in a post-hoc analysis, we categorize them into three groups: Group A represents no redundancy, Group B represents low redundancy, and Group C encompasses medium, high, and full redundancy. Blocks where the DS was not purchased are represented with a dash-dot-dash line (-.-); all other lines depict blocks where the DS was purchased.

Table 2. Results of GLMM Model for Blocks in Which Participants Purchased DS.

	Effective Sensitivity			Score		
	b	SE b	95% CI	b	SE b	95% CI
Baseline	.86***	.06	(.74, .97)	64.47***	1.84	(60.86, 68.07)
System d'	.7***	.03	(.63, .77)	21.77***	1.08	(19.65, 23.88)
Redundancy	.24**	.1	(.04, .44)	2.88	3.18	(-3.35, 9.12)
Block	.03***	.01	(.02, .06)	1.32***	.27	(.79, 1.85)
System d' X redundancy	-.48***	.06	(-.6, -.36)	-11.39***	1.93	(-15.17, -7.62)

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

system d' X redundancy ($p = .2$) were not significant. The analyses for the blocks where participants decided to purchase the DS showed significant effects for system d', redundancy, block, and the interaction system d' X redundancy (see Table 2).

To understand the interactions we conducted separate GLMMs for each level of system d' (with redundancy level as a fixed effect). The effect of redundancy was significant in all levels of system d'. We then recoded redundancy and system d' as factors and calculated contrasts between the various levels using the Tukey (HSD) test. To streamline the reporting of differences between redundancy levels, we categorize them into three groups: Group A represents no redundancy, Group B represents low redundancy, and Group C encompasses medium, high, and full redundancy, as can be seen in Figure 5. A was significantly different from B and C, $p < .001$, and B was different from C, $p < .001$. There were also significant differences between every pair of system sensitivity levels, $p < .001$. Thus, while higher levels of redundancy lower the effective sensitivity, higher levels of the system's d' increase it.

Separate contrast analyses for each level of DS sensitivity showed that for low DS sensitivity, the only significant difference was between no-redundancy and the other redundancy levels ($p < 0.01$). In medium DS sensitivity, all redundancy levels except between medium and high showed significant differences ($p < 0.01$). In high DS sensitivity, significant effects were observed between no-redundancy and all other redundancy levels and between low redundancy and others ($p < 0.001$). As shown in Figure 5, the impact of redundancy levels is more pronounced with high DS sensitivity.

Score. Results show similar patterns. There was a significant effect of PD, with significantly higher scores when DS was purchased ($M = 95.35$; $SD = 10.86$), compared to blocks where DS was not purchased ($M = 89.98$; $SD = 7.21$). Moreover, there was a learning effect, as the score increased as participants gained experience in the task, ($M_1 = 91.34$, $SD_1 = 10.45$; $M_2 = 93.57$, $SD_2 = 9.12$; $M_3 = 94.69$, $SD_3 = 9.86$). Table 2 and Figure 5 show the results of the model.

We then analyzed scores separately for participants who purchased (see Table 2 for the results) and did not purchase the DS. Among participants who decided not to purchase the DS, again, only the block was significant, $b = 1.93$, $p < .001$, $SE = .37$, 95% CI [1.19–2.66]. A post-hoc analysis using the Tukey HSD test revealed a significant difference between test block A ($M = 87.89$; $SD = 7.39$), test block B ($M = 90.53$; $SD = 6.37$), $p < .01$, and test block C ($M = 91.74$; $SD = 7.35$), $p < .001$. System d' ($p = .15$), redundancy level ($p = .42$), and interaction system d' X redundancy ($p = .26$) were not significant.

We applied the same GLMM analysis as used for the effective sensitivity to examine differences between levels of redundancy and system d' among participants who purchased DS, and found similar results.

Trust

Participants' subjective trust in the DS was measured for blocks DS was purchased with two items scaled between 1 and 10. The two trust items were correlated ($r = 0.49$, $p < .001$). Consequently, we calculated the mean trust score and conducted a 2-way Analysis of Variance (ANOVA) on the

combined score with redundancy level and system d' as independent variables. The main effects of redundancy level was significant, $F(4,812) = 40.03, p < .001, \eta^2 = .070$, as was the effect of system d' , $F(2,812) = 676.44, p < .001, \eta^2 = .55$, and the interaction redundancy level \times system d' , $F(2,812) = 16.51, p < .001, \eta^2 = .05$, see Figure 6, Table 3.

Separate analyses for each level of system sensitivity showed a significant effect of redundancy level for all three levels of system d' (for the system with low sensitivity, $F(4,169) = 11.81, p < .001, \eta^2 = .22$; for medium sensitivity, $F(4,305) = 37.84, p < .001, \eta^2 = .33$; and for high sensitivity, $F(4,332) = 26.33, p < .001, \eta^2 = .24$). However, the direction of the effects differed depending on the system sensitivity level, as can be seen in Figure 6. Participants accessing the low-sensitivity system reported the highest trust with the fully redundant DS. In contrast, participants with medium or high sensitivity systems reported the highest trust when the DS was nonredundant.

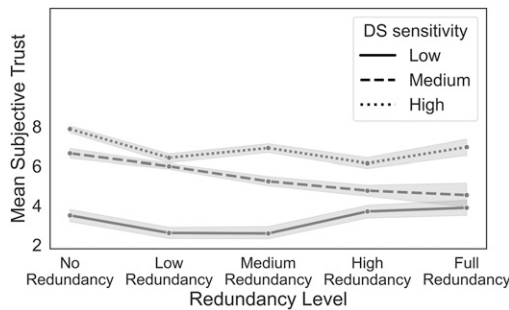


Figure 6. Mean trust in the DS by system d' and redundancy. The error ranges represent standard errors.

Discussion

We studied the choice and use of the information from Decision Support (DS) as a function of the level of redundancy and the quality of the available information. Normative models predict that with two (or more) equally good sources of information, decision makers should choose to receive nonredundant information, give it greater weight in decisions, and trust it more because it contains more unique information. In contrast, redundant information will repeat already available information and likely to lower the chances of contradictions between information from different sources.

Our results suggest that users tend to prefer less correlated information sources, regardless of the system's reliability, as is evident from the monotonic decrease in purchase decisions with increasing redundancy (Figure 4). This might indicate that people understand the value of nonredundant information (Goethals & Nelson, 1973), and they mostly adopt a normative approach (Gonzalez, 1994) rather than a confirmative one. However, some participants in our experiment used the more redundant and even fully correlated systems. This behavior aligns with findings by Kahneman and Tversky (1973, 1974), who demonstrated that people tend to have greater confidence in predictions derived from redundant input variables. Participants accessed the information from systems with higher sensitivity more, which replicates the results of previous studies (Rieger & Manzey, 2022). Results showed no effect of the interaction between the DS sensitivity and redundancy on the tendency to purchase DS.

Table 3. Means and (Standard Deviations) for the Trust in the Different Combinations of DS Sensitivity and Redundancy.

		Trust		
		DS Sensitivity Level		
		Low	Medium	High
DS redundancy level	No	3.5 (1.17)	6.63 (1)	7.86 (.91)
	Low	2.61 (.96)	5.99 (.79)	6.41 (1.02)
	Medium	2.58 (.92)	5.23 (.88)	6.9 (.94)
	High	3.7 (.95)	4.76 (1.16)	6.14 (1.09)
	Full	3.89 (.97)	5.53 (1.83)	6.95 (1.64)

A possible explanation for why participants in our experiment purchased the fully redundant system could be because it provides a binary indication, which is easier to interpret than numeric values like those presented as the participants' observed value (Fisher & Keil, 2018). Another explanation could be that people prefer to gather more information, even when it is worthless (Baron & Hershey, 1988), or gather information that confirms preexisting biases (Iyengar & Hahn, 2009).

Because the level of redundancy was varied as a between-subjects variable, each participant experienced only one type of DS. We did not expect the participants' underlying preference for redundancy/nonredundancy to change over the course of an experiment. However, the effect of encountering different sources of information can be investigated in future studies.

We assessed participant performance with the cumulative score and the effective sensitivity. As shown in Figure 5, in both measures participants performed better with nonredundant DS, in line with a previous study (Elvers & Elrif, 1997). This finding underscores the importance of examining the quality of the DS we interact with and its correlation with already available information, as it can significantly impact performance. The similar effect of redundancy on the two measures highlights the interconnectedness between decision outcomes and perceptual accuracy, indicating that participants who excel in rewards-based performance also exhibit an enhanced ability to discern meaningful information from noise, and vice versa.

Our results resemble those of previous studies in that (regardless of the redundancy level) DS mostly helped to improve performance (Bartlett & McCarley, 2017), and more valid information led to greater sensitivity (Bartlett & McCarley, 2017; Parasuraman & Riley, 1997), see Figure 5. Results demonstrated a significant interaction between DS sensitivity and redundancy on performance. Specifically, redundancy when the DS sensitivity was high led to a stronger decrease in performance compared to scenarios with lower sensitivity.

Trust is a major factor when using DS (Schaefer et al., 2016). As expected, participants expressed greater trust in the DS when it provided better information (i.e., higher sensitivity), replicating

previous studies (Rieger & Manzey, 2022). While Kahneman and Tversky (1973; 1974) indicate that redundant information increases confidence, results from our study showed a complex effect of redundancy on subjective trust (see Figure 6). With low sensitivity DS, participants reported the highest trust when the system provided redundant information, possibly due to a desire for consistency. Participants reported the highest trust with medium or high DS sensitivity when the DS was nonredundant. Trust here seems to reflect the value of the information for improving decision making. This finding reveals an interesting relationship between system redundancy and sensitivity, demonstrating that trust in DS is not merely about sensitivity alone or redundancy alone but depends on the interaction of the two. In environments where DS sensitivity is compromised, redundancy builds trust, and users might trust DS more when it provides redundant information. However, in contexts where systems have medium or high sensitivity, we should expect users to trust more on nonredundant DS.

Constraints on Generality

The study utilized an abstract task designed for nonexpert decision makers. Necessarily, it differs from many real-world tasks in ways that might limit the generalizability of our findings. For example, our experiment includes a specific set of parameters that may not accurately reflect real-world situations. Future research should explore the effects of different event distributions and classification costs, as well as more complex and less artificial tasks and designs. Moreover, the DS used in the experiment provides binary information. Thus, our findings might not generalize to interactions with more sophisticated systems that offer probabilistic or graded information. Another limitation is our between-subjects design, where participants were exposed to only one type of DS, preventing within-subject comparisons of different dependent variables.

Conclusion

Most analyses of automation-aided decision making assume that humans and automation have uncorrelated information. However, in actual systems, this assumption may not always be correct. We

examined how redundancy (i.e., correlated information), quality of the Decision Support (DS), and their interaction affected DS usage and users' evaluation of the system. Participants' responses were in line with the prescriptions of the normative model as they tended to purchase more uncorrelated, nonredundant information. They perform better with high-quality and nonredundant DS and trust the high-quality DS more. The interaction between DS sensitivity and redundancy significantly impacts both performance and trust, with high sensitivity amplifying the negative effects of redundancy on performance, while redundancy enhances trust in a less accurate system. Our study demonstrates that the evaluation of systems in the development and deployment stages should include careful analyses of the information needed to accomplish tasks and their availability from different sources, including assessments of the redundancy between information sources.

Key Points

- Participants generally preferred to use uncorrelated information and high-sensitivity automation, but there were still systematic deviations from optimal behavior.
- Task performance decreased as a function of the redundancy of the information from DS, with a higher decrease with high DS sensitivity, highlighting the need to carefully examine the performance of decision support systems and their correlation with available information.
- The perception of consistency in information is an important factor for trust in lower quality decision support, while trust reflects the informative value of higher quality decision support.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by ISF (Israel Science Foundation) grant 2029/19 to JM and was

partially supported by the Koret Foundation Grants for Smart Cities and Digital Living 2030 and by a grant from the Tel Aviv University Center for AI and Data Science (TAD).

ORCID iDs

Yoav Ben Yaakov  <https://orcid.org/0000-0001-9834-6568>

Maja Denisova  <https://orcid.org/0000-0001-8628-657X>

Joachim Meyer  <https://orcid.org/0000-0002-1801-9987>

References

- Alberdi, E., Povykalo, A., Strigini, L., & Ayton, P. (2004). Effects of incorrect Computer-Aided Detection (CAD) output on human decision-making in mammography. *Academic Radiology*, 11(8), 909–918. <https://doi.org/10.1016/j.acra.2004.05.012>
- Baron, J. & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: I. Priors, error costs, and test accuracy. *Organizational Behavior and Human Decision Processes*, 41(2), 259–279. [https://doi.org/10.1016/0749-5978\(88\)90030-1](https://doi.org/10.1016/0749-5978(88)90030-1)
- Bartlett, M. L. & McCarley, J. S. (2017). Benchmarking aided decision making in a signal detection task. *Human Factors*, 59(6), 881–900. <https://doi.org/10.1177/0018720817700258>
- Ben-Yaakov, Y., Bitan, Y., & Meyer, J. (2021). Do users know when to buy decision support and how to use it? *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting*, 65(1), 1021. <https://doi.org/10.1177/1071181321651120>
- Ben Yaakov, Y., McCarley, J. S., & Meyer, J. (2024). Selective access to decision support as a function of event uncertainty. *Human Factors*, 187208241277158. <https://doi.org/10.1177/00187208241277158>
- Ben Yaakov, Y. & Meyer, J. (2024). Choose and use: Users' selection of information sources for decision support. In *Proceedings of the annual meeting of the cognitive science society* (46). Cognitive Science Society.
- Ben-Yaakov, Y., Meyer, J., Wang, X., & An, B. (2020). User detection of threats with different security measures. *2020 IEEE International Conference on Human-Machine Systems (ICHMS)* (pp. 1–6). IEEE.
- Ben Yaakov, Y., Wang, X., Meyer, J., & An, B. (2019). Choosing protection: User investments in security measures for cyber risk management. *International*

- conference on decision and game theory for security (pp. 33–44). Springer-Verlag.
- Bonaccio, S. & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127–151. <https://doi.org/10.1016/j.obhdp.2006.07.001>
- Botzer, A., Meyer, J., & Parmet, Y. (2013). Mental effort in binary categorization aided by binary cues. *Journal of Experimental Psychology: Applied*, 19(1), 39–54. <https://doi.org/10.1037/a0031625>
- Cover, T. M. & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). John Wiley & Sons, Incorporated. <https://doi.org/10.1002/047174882X>
- Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*, 49(4), 564–572. <https://doi.org/10.1518/001872007X215656>
- Elvers, G. C. & Elrif, P. (1997). The effects of correlation and response bias in alerted monitor displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(4), 570–580. <https://doi.org/10.1518/001872097778667960>
- Fisher, M. & Keil, F. C. (2018). The binary bias: A systematic distortion in the integration of information. *Psychological Science*, 29(11), 1846–1858. <https://doi.org/10.1177/0956797618792256>
- Goethals, G. R. & Nelson, R. E. (1973). Similarity in the influence process: The belief-value distinction. *Journal of Personality and Social Psychology*, 25(1), 117–122. <https://doi.org/10.1037/h0034266>
- Gonzalez, R. (1994). When words speak louder than actions: Another's evaluations can appear more diagnostic than their decisions. *Organizational Behavior and Human Decision Processes*, 58(2), 214–245. <https://doi.org/10.1006/obhd.1994.1035>
- Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.
- Gupta, N., Bhatele, P., & Khanna, P. (2018). Identification of Gliomas from brain MRI through adaptive segmentation and run length of centralized patterns. *Journal of Computational Science*, 25, 213–220. <https://doi.org/10.1016/j.jocs.2017.02.009>
- Hogarth, R. M. (1989). On combining diagnostic 'forecasts': Thoughts and some evidence. *International Journal of Forecasting*, 5(4), 593–597. [https://doi.org/10.1016/0169-2070\(89\)90015-0](https://doi.org/10.1016/0169-2070(89)90015-0)
- Iyengar, S. & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1), 19–39. <https://doi.org/10.1111/j.1460-2466.2008.01402.x>
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237–251. <https://doi.org/10.1037/h0034747>
- Kaiser, H. F. & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika*, 27(2), 179–182. <https://doi.org/10.1007/bf02289635>
- Kraus, M. & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, 104, 38–48. <https://doi.org/10.1016/j.dss.2017.10.001>
- Lee, J. D. & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Macmillan, N. A. & Creelman, C. D. (2004). Detection theory: A user's guide. *Psychology press*, 136(Issue 2).
- Maines, L. A. (1990). The effect of forecast redundancy on judgments of a consensus forecast's expected accuracy. *Journal of Accounting Research*, 28, 29–47. <https://doi.org/10.2307/2491245>
- Maltz, M. & Meyer, J. (2001). Use of warnings in an attentionally demanding detection task. *Human Factors*, 43(2), 217–226. <https://doi.org/10.1518/001872001775900931>
- Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors*, 43(4), 563–572. <https://doi.org/10.1518/001872001775870395>
- Meyer, J. & Kuchar, J. K. (2021). Maximal benefits and possible detrimental effects of binary decision aids. *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)* (pp. 1–6). IEEE.
- Mulugeta, F., Ben Yaakov, Y., & Meyer, J. (2022). Confirmative or normative? Information redundancy in decision support systems. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting*, 66(1), 2042–2046. <https://doi.org/10.1177/1071181322661121>
- Munoz Gomez Andrade, F., Duncan-Reid, J., & McCarley, J. S. (2022). The effect of correlated observations on human-automation interaction in a signal detection task. *Proceedings of the Human*

- Factors and Ergonomics Society - Annual Meeting*, 66(1), 2062–2066. <https://doi.org/10.1177/1071181322661290>
- Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, 2(2), 140–160. <https://doi.org/10.1518/155534308x284417>
- Parasuraman, R. & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors*, 50(3), 511–520. <https://doi.org/10.1518/001872008X312198>
- Pollack, I. & Madans, A. B. (1964). On the performance of a combination of detectors. *Human Factors*, 6(5), 523–531. <https://doi.org/10.1177/001872086400600510>
- Rieger, T. & Manzey, D. (2022). Human performance consequences of automated decision aids: The impact of time pressure. *Human Factors*, 64(4), 617–634. <https://doi.org/10.1177/0018720820965019>
- Robinson, D. E. & Sorkin, R. D. (1985). A contingent criterion model of computer assisted detection. *Trends in ergonomics/human factors*, 2, 75–82.
- Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, 58(3), 377–400. <https://doi.org/10.1177/0018720816634228>
- Seabold, S. & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *SciPy*, 7(1).
- Soll, J. B. (1999). Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psychology*, 38(2), 317–346. <https://doi.org/10.1006/cogp.1998.0699>
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Yaniv, I. (2004). The benefit of additional opinions. *Current Directions in Psychological Science*, 13(2), 75–78. <https://doi.org/10.1111/j.0963-7214.2004.00278.x>

Author Biographies

Yoav Ben Yaakov is a PhD candidate in the Department of Industrial Engineering at Tel Aviv University, Israel. He has an MSc degree in industrial engineering (2020) from Tel Aviv University.

Maja Denisova is a graduate of the MSc Human Factors program at Technische Universität Berlin, Germany. In her exchange semester, she studied at Tel Aviv University, followed by a research internship in JM's lab at the Department of Industrial Engineering, Tel Aviv University.

Filmona Mulugeta is a graduate of the MSc Human Factors program at Technische Universität Berlin, Germany. In her exchange semester, she studied at Tel Aviv University at the Department of Industrial Engineering, Tel Aviv University.

Joachim Meyer is the Celia and Marcos Maus Professor of Data Sciences at the Department of Industrial Engineering at Tel Aviv University, Israel. He was on the faculty of Ben-Gurion University of the Negev, Israel. He held research positions at the Technion, Israel Institute of Technology, MIT AgeLab, and MIT MediaLab. He has an MA degree in psychology and a PhD degree in industrial engineering (1994) from Ben-Gurion University of the Negev in Beer Sheva, Israel.