Behringer, S., Belavkin, R.V. (2026). *Value of Information in Bayesian Environments*. In: Moosaei, H., Belavkin, R., Pardalos, P.M. (eds) *Dynamics of Information Systems: DIS 2025*. Lecture Notes in Computer Science, vol 15940. Springer, Cham.

The Version of Record is available online at:
https://doi.org/10.1007/978-3-032-16469-8_2

Value of Information in Bayesian Environments

Stefan Behringer[1][0000−0002−7026−5149] and Roman V. Belavkin[2][0000−0002−2356−1447]

[1] Universität Bielefeld, Value of Information Group, Germany
stefan@stefanbehringer.com
https://www.stefanbehringer.com
[2] Department of Computer Science, Middlesex University, London, NW4 4BT, UK
R.Belavkin@mdx.ac.uk
https://www.mdx.ac.uk/about-us/our-people/staff-directory/dr-roman-belavkin/

Abstract. The Value of Information (VoI) framework, developed by Ruslan Stratonovich, bridges Claude Shannon's information theory with economics, particularly utility and decision theory. This paper revisits the VoI concept in the Boolean utility setting of hypothesis testing and then extends the discrete Bayesian VoI framework to more general decision contexts with arbitrary utility matrices. We also provide a method to compute a *robust* VoI estimate that is independent of input distributions.

Keywords: Value of Information · Shannon's Information · Bayesian Systems · Robust VoI · Economics · Circulant Matrices

1 Introduction

The *Value of Information* (VoI) framework builds on foundational contributions from Ruslan Stratonovich and Claude Shannon. Originally developed within the settings of rate distortion theory and statistical mechanics, VoI has gained renewed interest due to advancements in data science, machine learning, economics, and statistical inference. Shannon's information theory quantifies the reduction of uncertainty, while Stratonovich extended it by linking information directly to decision-making, thereby connecting it to economic risk. A key insight is that better information reduces risk in a way that aligns with sub-additive risk measures.

For economics, VoI is fundamental because it bridges information theory and its concepts with traditional economics (e.g. utility theory, stochastic finance, and decision theory). It allows for continuous modelling of uncertainty and risk while maintaining the optimization paradigm central to traditional economics. An extension of the classical VoI framework to economic utility functions is provided in [5], showing how the classical quadratic-loss, variance-based settings can be generalized to accommodate more general risk measures beyond variance. VoI in Zero-sum games is analysed in [2], while a comprehensive treatment of the VoI framework, including continuous settings and applications, is presented in the forthcoming book [3]. In this paper, we focus to discrete Bayesian settings.

2 Value of Information

Following Ruslan Stratonovich [4, 8, 9], the Value of Information (VoI) is a function defined as

$$V(I) = U(I) - U(I_{\min})$$

for *Shannon mutual information* $I(X, Z)$ where

$$U(I) \equiv \sup[E_\omega\{u(x, z)\} : I(X, Z) \leq I]$$

or equivalently

$$I(U) \equiv \inf[I(X, Z) : E_\omega\{u(x, z)\} \geq U]$$

by strong duality, where the conditional extremum is taken over all joint probability measures ω over a system $X \times Z$. Measurable functions $u : X \times Z \to \mathbb{R}$ are interpreted as *economic utility functions* (or changing the sign) as cost functions. $U(I_{\min})$ may often be treated as a constant. With a minimum of (often simply no) information about X we just have

$$U(I_{\min}) \equiv \sup_{z \in Z}[E_{P(x)}\{u(x, z)].$$

Often there may be another random variable (a message or signal) Y that communicates information about X. If it communicates *full* information then there is an invertible function $y = f(x)$ such that $x = f^{-1}(y)$ is determined uniquely by the message $y \in Y$ and the optimal value can be found as

$$U(\infty) \equiv \sup_{z(y)}[E_{P(x)}\{u(x, z(y))\}$$

where optimizations takes place over all mappings $z(y)$, i.e. $z : Y \to Z$.

There are different ways in which the information amount I and quantity $U(I)$ can be defined. Stratonovich also presents the concepts of Hartley's and Boltzmann's VoI. The main result of Stratonovich [9] is their asymptotic equivalence.

Recall that Shannon mutual information is defined as

$$I(X, Z) \equiv H(X) - H(X \mid Z) \leq \min[H(X), H(Z)].$$

For less than full information we thus have the non-constant component of Stratonovich VoI as

$$U(I) \equiv \sup_{P(z|x)}[E_{P(x,z)}\{u(x, z)\} : I(X, Z) \leq I]$$

where

$$P(z \mid x) = \sum_Y P(z \mid y)P(y \mid x) \text{ and } I(X, Y) \leq I.$$

We next investigate the use of VoI in settings that allow for simple Boolean utility.

2.1 VoI calculations for Boolean utility

Boolean utility refers to a decision-making or evaluation approach where the utility of an outcome is represented using binary values that are typically just false/true or 0/1. This is common in scenarios where an outcome is either acceptable or not, rather than having a graded or probabilistic utility.

Joint measures that are optimal by Lagrange multiplier methods are:

$$W(x, z; \beta) = P(x)Q(z)e^{\beta u(x,z)-\gamma(\beta,x)}$$

see e.g. [4], where $P(x)$ and $Q(z)$ are the marginals of W, β^{-1} is the *temperature*, the Lagrange multiplier (shadow price) of the constraint $I(X, Z) \leq I$ and $\gamma(\beta, x)$ is defined by the normalization $\sum_{x,z} W(x, z; \beta) = 1$.

Using partial traces we find the conditions

$$\sum_x W(x, z) = Q(z) \Rightarrow \sum_x e^{\beta u(x,z)-\gamma(\beta,x)} P(x) = 1$$

$$\sum_z W(x, z) = P(x) \Rightarrow \sum_z e^{\beta u(x,z)} Q(z) = e^{\gamma(\beta,x)}.$$

VoI in hypothesis testing has been investigated in [6] before. It implies a two-state Bayesian system that is derived by avoiding type 1 error (rejecting the correct H_0 hypothesis (*not* $z_1 = z_2 = x_1$, i.e. false positives) and type 2 error (failing to reject the H_0 hypothesis when it is actually false $z_1 = x_2$).

Thus one finds a 2×2 action-outcome utility matrix for a hypothesis test/prediction z and data x as

$$u(x, z) = \begin{bmatrix} u(x_1, z_1) & u(x_1, z_2) \\ u(x_2, z_1) & u(x_2, z_2) \end{bmatrix} = \begin{bmatrix} u_{11} & \textit{type 1 error} \\ \textit{type 2 error} & u_{22} \end{bmatrix}.$$

Assuming a prior $p = P(x_1)$, in the absence of any information/data (y) the output distribution $Q(z)$ is a δ-distribution

$$Q(z_1) = \left\{ \begin{matrix} 1 \text{ if } E_p\{u \mid z_1\} \geq E_p\{u \mid z_2\} \Leftrightarrow \\ pu_{11} + (1-p)u_{21} \geq pu_{12} + (1-p)u_{22} \Leftrightarrow \\ p(u_{11} - u_{12}) + (1-p)(u_{12} - u_{22}) \\ 0 \text{ else} \end{matrix} \right\}.$$

So similar to the case of an econometric regression where we may have $-u = |z_1 - x_1|$ (implying *translation invariance* which will be investigated below) for some estimator z_1 here z_1 is a hypothesis and so the *confusion matrix* may simply be

$$u(x, z) = \begin{bmatrix} u(x_1, z_1) & u(x_1, z_2) \\ u(x_2, z_1) & u(x_2, z_2) \end{bmatrix} = \begin{bmatrix} c_1 + d_1 & c_1 - d_1 \\ c_2 - d_2 & c_2 + d_2 \end{bmatrix}.$$

Following [6] letting $c_1 = c_2 = \frac{1}{2}$ and $d_1 = d_2 = \frac{1}{2}$ and equal priors we can derive

$$I(U) = \ln 2 + U \ln U + (1 - U)\ln(1 - U).$$

which depicts the Value of Information $V(I) = U(I) - U(0)$ in this Boolean setting. Note that this choice of parameters renders the utility matrix into the identity matrix which makes the VoI independent of the input distribution and hence *robust*. However an identity utility matrix (being also trivially Boolean), is sufficient but not necessary to the robustness result. In fact it is necessary that underlying utility is *translation invariant*.

Theorem 1. Translation invariance *of the utility function is necessary and sufficient for the action-outcome (utility/cost) matrix in the discrete setting to be* circulant*.*

Proof. Elementary, see e.g. [3]. ■

3 Beyond Boolean Utility

A first extension to Boolean utility is a 3×3 *action-outcome matrix* with three possible utility values.

Example 1. The Rock-Paper-Scissors game also has a circulant matrix

$$\begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$$

and

$$U(\beta) = \frac{2\sinh(\beta)}{1 + 2\cosh(\beta)}.$$

The VoI results $V(I)$ will not depend on the input distribution, by Theorem 1 and thus give *robust* predictions.

However, in most practical cases utility matrices that are exactly circulant will be the exception. How, then, can we proceed in such cases?

Example 2. This 3×3 Magic Square has a non-circulant utility matrix

$$\begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix}$$

and the resulting VoI $V(I)$ will depend on the input distribution. Assuming a uniform prior the VoI can be derived parametrically (see Figure 1) but this result is not robust. Whence we employ the following theorem:

Theorem 2. *Using the Discrete Fourier Transform (DFT), any non-circulant generic matrix can be expressed exactly as a sum of rank-one terms in the Fourier basis, and can be approximated by a circulant matrix obtained by retaining only the diagonal DFT coefficients.*[3]

Proof. See the following 3×3 Example: For $A \in \mathbb{C}^{3\times 3}$, let us illustrate the DFT decomposition with $N = 3$. The unitary Fourier vectors are

$$v_0 = \tfrac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v_1 = \tfrac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ \omega \\ \omega^2 \end{pmatrix}, \quad v_2 = \tfrac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ \omega^2 \\ \omega \end{pmatrix},$$

where $\omega = e^{-2\pi i/3}$ is a primitive third root of unity. Thus the unitary Fourier matrix is

$$F = \tfrac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix}, \qquad F^* = F^{-1}.$$

The Fourier transform of A is

$$\widehat{A} \;=\; FAF^* \;\in\; \mathbb{C}^{3\times 3}.$$

If A is circulant, then F is the DFT matrix, which forms the eigenbasis of A, and $\widehat{A}$ is the diagonal matrix of its eigenvalues. The full spectral expansion of any matrix A (circulant or not) in the DFT eigenbasis is

$$A \;=\; \sum_{k,\ell=0}^{2} \widehat{A}_{k\ell}\, v_k v_\ell^*.$$

Each term $v_k v_\ell^*$ is a rank-one circulant-like matrix. A star* denotes the conjugate transpose, so that

$$v_k v_\ell^* = \begin{pmatrix} (v_k)_1 \\ (v_k)_2 \\ (v_k)_3 \end{pmatrix} \begin{pmatrix} \overline{(v_\ell)_1} & \overline{(v_\ell)_2} & \overline{(v_\ell)_3} \end{pmatrix},$$

whose (i,j)-entry is $(v_k)_i\, \overline{(v_\ell)_j}$. We thus find:

$$v_0 v_0^* = \tfrac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad v_1 v_1^* = \tfrac{1}{3} \begin{pmatrix} 1 & \overline{\omega} & \overline{\omega^2} \\ \omega & 1 & \overline{\omega} \\ \omega^2 & \omega & 1 \end{pmatrix}, \qquad v_2 v_2^* = \tfrac{1}{3} \begin{pmatrix} 1 & \overline{\omega^2} & \overline{\omega} \\ \omega^2 & 1 & \overline{\omega^2} \\ \omega & \omega^2 & 1 \end{pmatrix},$$

[3] Note this Theorem reverses the classical focus where DFT is the eigenbasis that diagonalizes all circulant matrices, as e.g. in [1], as our method focuses on employing the DFT to project non-circulant matrices onto the space of circulant matrices. In classical DFT applications (signal processing, convolution, filtering), the focus is almost always on circulant matrices.

where the overline on $\overline{\omega}$ denotes convex conjugates. The diagonal terms are explicitly circulant when the conjugates are rewritten as:

$$v_0 v_0^* = \tfrac{1}{3}\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix}, \quad v_1 v_1^* = \tfrac{1}{3}\begin{pmatrix}1&\omega^2&\omega\\ \omega&1&\omega^2\\ \omega^2&\omega&1\end{pmatrix}, \quad v_2 v_2^* = \tfrac{1}{3}\begin{pmatrix}1&\omega&\omega^2\\ \omega^2&1&\omega\\ \omega&\omega^2&1\end{pmatrix}.$$

Hence, grouping terms according to Fourier modes, we can exactly express

$$A = \underbrace{\sum_{k=0}^{2} \widehat{A}_{kk}\, v_k v_k^*}_{\text{diagonal (DC) part}} + \underbrace{\sum_{k\neq\ell} \widehat{A}_{k\ell}\, v_k v_\ell^*}_{\text{off-diagonal Fourier modes}}$$

where each term $v_k v_k^*$ and $v_k v_\ell^*$ is a rank-one matrix in the Fourier basis. Whence the sum of the diagonal terms in the DFT expansion can be written explicitly as

$$\sum_{k=0}^{2} \widehat{A}_{kk}\, v_k v_k^* = \frac{1}{3}\left[\widehat{A}_{00}\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix} + \widehat{A}_{11}\begin{pmatrix}1&\omega^2&\omega\\ \omega&1&\omega^2\\ \omega^2&\omega&1\end{pmatrix} + \widehat{A}_{22}\begin{pmatrix}1&\omega&\omega^2\\ \omega^2&1&\omega\\ \omega&\omega^2&1\end{pmatrix}\right].$$

Similarly, the off-diagonal terms can be written explicitly as

$$\sum_{k\neq\ell} \widehat{A}_{k\ell}\, v_k v_\ell^* = \frac{1}{3}\left[\widehat{A}_{01}\begin{pmatrix}1&\omega^2&\omega\\1&\omega^2&\omega\\1&\omega^2&\omega\end{pmatrix} + \widehat{A}_{02}\begin{pmatrix}1&\omega&\omega^2\\1&\omega&\omega^2\\1&\omega&\omega^2\end{pmatrix} + \dots\right].$$

Each $v_k v_\ell^*$ is a rank-one matrix, however the diagonal terms $v_k v_k^*$ are circulant, while the off-diagonal terms are structured but need not be circulant.

Whence using the DFT, any generic non-circulant matrix (square, finite dimensional) can be approximated as a sum of circulant matrices by keeping only the diagonal Fourier terms. ■

Employing the algorithm provided in [7] we can use DFT to decompose the Magic Square Matrix Example 2 exactly in the form $A = M_0D_0 + M_1D_1 + M_2D_2$:

$$\begin{bmatrix}8&1&6\\3&5&7\\4&9&2\end{bmatrix} = \begin{bmatrix}5&4&6\\6&5&4\\4&6&5\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} + \begin{bmatrix}1.5-0.86i & 1.73i & -1.5-0.86i\\ -1.5-0.86i & 1.5-0.86i & 1.73i\\ 1.73i & -1.5-0.86i & 1.5-0.86i\end{bmatrix}\begin{bmatrix}1&0&0\\0&e^{i\frac{2\pi}{3}}&0\\0&0&e^{i\frac{4\pi}{3}}\end{bmatrix} +$$

$$\begin{bmatrix}1.5+0.86i & -1.73i & -1.5+0.86i\\ -1.5+0.86i & 1.5+0.86i & -1.73i\\ -1.73i & -1.5+0.86i & 1.5+0.86i\end{bmatrix}\begin{bmatrix}1&0&0\\0&e^{i\frac{4\pi}{3}}&0\\0&0&e^{i\frac{2\pi}{3}}\end{bmatrix}$$

To calculate the VoI in a discrete setting for any generic non-circulant utility matrix, one may then proceed by the following *method*:

1. Form a low-frequency circulant approximation U_{approx} by retaining the DC term and, if doing so still preserves circulancy. One may add harmonics if that also preserves circulancy. Adding harmonics improves the approximation with no downside if circulancy is maintained.

2. Evaluate $\text{VoI}(U_{\text{approx}})$. The computation is straightforward because U_{approx} is circulant, which decouples the variational problem.

3. Treat $\text{VoI}(U_{\text{approx}})$ as a low-frequency approximation to $\text{VoI}(U)$. This gives a tractable estimate capturing the dominant contribution to the VoI while ensuring prior-independence and robustness.

Only the real parts of the matrices are used because utility must be real. The expected payoff $E\{u(X,Y)\}$ is a real scalar, which makes decision-theoretic comparisons meaningful. Thus the utility in optimization is taken as $u = \Re h$. Likewise, information measures such as mutual information and entropy are real and non-negative, so the VoI is also real. Any complex values appearing in intermediate steps cancel or reduce to real values in the final expected utility and VoI calculations.

See from the decomposition that $M_1D_1 + M_2D_2 = 2\Re(M_1D_1)$ is circulant, so one may add the full first harmonic which is equivalent to twice the real part of $\Re(M_1D_1)$. So a good approximation of the utility matrix is obtained by keeping the DC term and, in this case, also the full first harmonic:

$$\text{DC} = \begin{pmatrix} 5 & 4 & 6 \\ 6 & 5 & 4 \\ 4 & 6 & 5 \end{pmatrix}, \quad \text{first harmonic (real part)} = 2 * \begin{pmatrix} 1.5 & 0 & -1.5 \\ -1.5 & 1.5 & 0 \\ 0 & -1.5 & 1.5 \end{pmatrix}.$$

Adding these matrices gives a circulant approximation of the original utility:

$$U_{\text{approx}} = \begin{pmatrix} 5+3 & 4 & 6-3 \\ 6-3 & 5+3 & 4 \\ 4 & 6-3 & 5+3 \end{pmatrix} = \begin{pmatrix} 8 & 4 & 3 \\ 3 & 8 & 4 \\ 4 & 3 & 8 \end{pmatrix}.$$

Retaining the full first harmonic, the resulting U_{approx} is both real and circulant, which allows us to compute the VoI as for a single real matrix *without requiring any assumption on the prior distribution.* By Theorem 2, this approach generalizes to any generic $N \times N$ matrix.

We obtain the *robust* Value of Information of the Magic Square from a parametric plot in Figure 1 below.

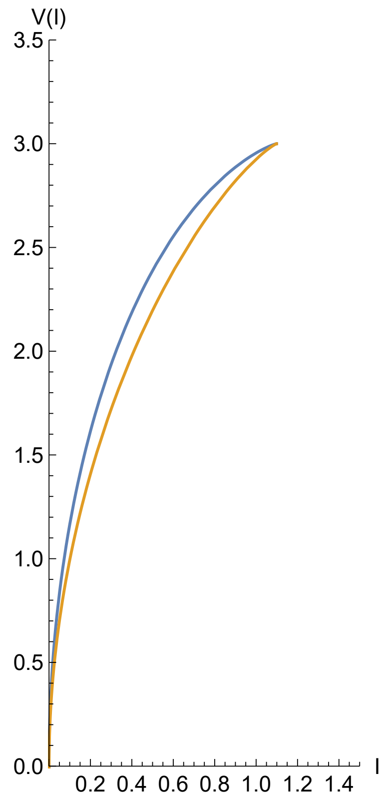

Fig. 1. VoI for a uniform Magic Square (blue) and *robust* VoI (yellow).

4 Conclusion

In this work we have shown how the Discrete Fourier Transform (DFT) provides a natural framework for analysing arbitrary utility matrices in the calculation of the Value of Information (VoI). Any finite $N \times N$ matrix can be decomposed into a sum of rank-one Fourier components. Among these, the DC component (the average utility) is always real and circulant, and therefore directly yields a simple, robust approximation. This approximation is particularly valuable because circulant matrices lead to decoupled variational problems, so that $\mathrm{VoI}(U_{\mathrm{approx}})$ can be computed independently of the prior distribution. In this sense, the DC approximation already provides a prior-independent benchmark for decision-making.

A more refined approximation is sometimes possible by selectively including the harmonics. In general the first harmonic arises as a conjugate pair of Fourier modes and both must be combined to recover the full contribution. If one component turns out to be circulant so is the other so we can include the full first harmonics in the approximation as in Example 2. The resulting U_{approx}

is a better approximation of the original matrix while maintaining the robust, prior-free structure that makes the analysis tractable. This selective inclusion reflects a balance: one seeks to capture as much of the original utility structure as possible without breaking circulancy. More generally, for higher-dimensional matrices the same principle w.r.t. higher harmonics applies.

The guiding rule is therefore: keep the DC component, and optionally add harmonics if their real parts remain circulant. This procedure ensures a systematic and robust way to approximate VoI in discrete settings, capturing dominant low-frequency structure while retaining tractability and independence from the input distribution. A more complete treatment of VoI in economics, decision theory, and data science can be found in [3].

Acknowledgments. We are grateful to Felix Otto for initial encouragement, Hariprasad Manjunath for the MATLAB code, participants of DIS 2025 London, the editor, and three anonymous referees for comments. This study received no funding.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bamieh, B.: Discovering transforms: A tutorial on circulant matrices, circular convolution, and the discrete fourier transform, arXiv preprint (2018), arXiv:1805.05533 https://arxiv.org/abs/1805.05533.
2. Behringer, S.: Value of Information in Zero-Sum Games. MaxEnt23 Presentation, mimeo (2023).
3. Behringer, S.: The Value of Information in Economics and Decision Theory, Springer (in preparation), (2026).
4. Behringer, S. and Belavkin, R.V.: The Value of Information and Circular Settings. arXiv preprint (2023), https://doi.org/10.48550/arXiv.2303.16126
5. Behringer, S. and Belavkin, R.V.: The Value of Information in Economic Contexts, Physical Sciences Forum (forthcoming) (2025).
6. Belavkin, R.V., Pardalos, P., and Principle J.: Value of Information in the Binary Case and Confusion Matrix. Physical Sciences Forum, (2022). https://doi.org/10.3390/psf2022005008
7. Manjunath, H. and Venkatapathi, M.: Circulant decomposition of a matrix and the eigenvalues of Toeplitz type matrices. arXiv (2022). https://arxiv.org/abs/2105.14805
8. Stratonovich, R.L.: On the Value of Information. Izv. USSR Acad. Sci. Tech. Cybern., Vol. 5. pp. 2-12 (1965).
9. Stratovonich, R.L.: Theory of Information and its Value. ed. R.V. Belavkin, R.M. Pardalos, J.C. Principe, Springer, Heidelberg (2020). https://doi.org/10.1007/978-3-030-22833-0