AMERICAN PSYCHOLOGICAL ASSOCIATION

Decision

ISSN: 2325-9965

2024, Vol. 11, No. 4, 684–699
https://doi.org/10.1037/dec0000248

Decisions as Ill-Posed Problems: A Scoping Review of Regularization Methods in Decision Science

Janina A. Hoffmann
Department of Psychology, University of Bath

Real-life decisions often require individuals to make inferences based upon a large number of possible predictive features, but only limited observations. These so called ill-posed problems can be translated into well-structured ones using regularization methods. While regularization methods using penalty terms have gained popularity in machine learning, decision science has increasingly adopted Bayesian approaches to regularization, such as hierarchical Bayesian estimation. Yet, theories of human decision making appear to have less often embraced the original framing of regularization methods as "solutions to ill-posed decision problems." In a scoping review, I investigate how regularization methods have influenced decision science in applications, contributed to methodological advances, and informed theoretical debates. Within the 117 reviewed articles, most studies applied regularization methods as a tool to deal with heterogeneity and nested data structures, to select the most relevant problem features, or to improve interpretability. Methodological advances through regularization have rendered cognitive choice models testable on high-dimensional data or allowed to map descriptive models onto normative models of human decision making. Theoretical innovations were, however, confined to modeling constraints in information processing, re-integrating heuristic with rational choice models, and debating the role of prior beliefs. Looking ahead, I discuss how future applications in decision may harness regularization techniques to communicate insights from the analysis of decision policies, to improve forecasting tools, and to innovate theoretical approaches in decision science.

Keywords: regularization, decision making, hierarchical Bayes, informative prior

Supplemental materials: https://doi.org/10.1037/dec0000248.supp

Which personality traits do you appreciate most in a partner? Which amenities in a house are worth making a higher offer? In statistics, these classical human decisions are regarded as ill-posed, sparse problems with a high dimensionality. When identifying the most valuable traits in a partner, for instance, each potential partner may be characterized by a myriad of attitudes, traits, and habits,

This article was published Online First August 29, 2024.
Janina A. Hoffmann https://orcid.org/0000-0002-6246-2724
Data and material associated with the article are made available via the Open Science Framework at https://osf.io/dvp37/ (Hoffmann, 2024).

Open Access funding provided by University of Bath:

Janina A. Hoffmann played a lead role in conceptualization, data curation, methodology, visualization, writing–original draft, and writing–review and editing.

The data are available at https://osf.io/dvp37/.

The experimental materials are available at https://osf.io/dvp37/.

Correspondence concerning this article should be addressed to Janina A. Hoffmann, Department of Psychology, University of Bath, Bath BA2 7AY, United Kingdom. Email: j.a.hoffmann@bath.ac.uk

rendering it a high-dimensional problem. Some characteristics are likeable, others less so, but usually one cannot observe all of these characteristics on a single date, not even on multiple dates. Thus, we only possess sparse information about each partner. Finally, we only date a limited number of partners, or cases, and have to infer from this limited number of cases which characteristics we perceive as likeable. Typical linear regression models cannot handle such problems because the number of dimensions exceeds the number of cases and allows for multiple solutions, making it an ill-posed problem. In his seminal dissertation, Tikhonov (1963) suggested to solve such problems via regularization techniques (or methods) that find stable solutions for parameter estimates despite noisy measurements. Regularization techniques have been successfully applied to solving high-dimensional problems across diverse fields from vision (Poggio et al., 1985) to medicine (Castiglioni et al., 2021; Friedrich et al., 2023) and are considered a well-known standard approach to improve generalization, the ability to make accurate predictions for unseen examples, in machine learning (Hastie et al., 2016; Tian & Zhang, 2022), neural networks (Bejani & Ghatee, 2021; Girosi et al., 1995; Santos & Papa, 2022), and deep learning (Moradi et al., 2020).

Provided that many everyday decision problems are ill-posed, it is interesting to note that the initial formulation of regularization methods as "solutions to ill-posed problems" appears to have received limited attention within the field of decision science. This article synthesizes insights from previous research employing regularization methods in decision science with the goal to carve out theoretical connections between the two fields. After an introduction into regularization methods, a scoping review assesses how methodology, applications, and theoretical approaches in decision science have build upon regularization techniques. Next, I explore future avenues for integrating theories in judgment and decision making with regularization methods (and vice versa). I discuss limitations and the relationship to state-of-the-art theories in decision science.

An Introduction Into Regularization Methods

Regularization methods aim to turn ill-posed problems into well-posed problems and to find reasonable, stable parameter estimates despite unstable settings (Bickel et al., 2006). Such unstable settings may arise, for instance, when attempting to fit complex models with a large number of parameters to high-dimensional data, as is often the case in machine learning or in neural networks. Regularization achieves more stable parameter estimates by introducing a regularization term $\gamma\text{Pen}(w)$ into the overall loss function, $L(w, \gamma)$.

$$L(w, \gamma) = L_{\text{Model}}(w) + \gamma\text{Pen}(w). \qquad (1)$$

The model loss, $L_{\text{Model}}(w)$, accounts for the lack of goodness of fit to the data. The regularization term penalizes the model for irregular or nonsmooth solutions where the regularization parameter γ balances the trade-off between model fit and smoothness. Regularization methods thus attempt to find the best-fitting model under the constraint that model parameters are sufficiently regular (Poggio et al., 1985). For this reason, regularization offers a straightforward solution to the bias–variance trade-off (Geman et al., 1992; James et al., 2013)—the problem that too simple models with few parameters miss out on explaining important variance in the data, whereas too complex models explain random noise and fail to predict new observations.

In a classical linear regression model, for instance, the goodness of fit of the regression model across all cases i depends upon how well the observed dependant variable y_i can be predicted linearly from the observed predictor variables x_{ij} weighted by their importance, that is the regression coefficients β_j, where j indexes different predictors.

$$L(\beta) = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2. \qquad (2)$$

In a penalized linear regression model, the regression coefficients β minimize the residual sum of squares while penalizing for the distance between the regression coefficients and zero

$$L(\beta, \gamma) = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \gamma\|\beta\|_K^K, \qquad (3)$$

where K indicates the norm applied as a penalty. With a higher γ, model fit looses importance and the regression coefficients are more strongly dragged toward zero. Two norms are most often applied to penalize irregular regression coefficients:

Lasso regression applies the l_1-norm, $K = 1$, and penalizes the coefficients with $\|\beta\|_1^1 = \sum_{j=1}^{p} |\beta_j|$. As a result, all coefficients are diminished by a constant amount and smaller coefficients are set to 0. Ridge regression applies the l_2-norm, $K = 2$, and penalizes the coefficients with $\|\beta\|_2^2 = \sum_{j=1}^{p} \beta_j^2$. As a result, all coefficients are scaled proportionally by a constant factor, but not set to zero. Taken together, the Lasso regression performs feature selection and considers the features with the smallest coefficients as irrelevant. Ridge regression outperforms the Lasso regression when many, equal-sized predictors should be taken into account and multicollinearity needs to be controlled (McDonald, 2009). Both regularization methods prevent linear regression models from overfitting when the variance in the nonregularized coefficients is high, for instance, when the number of observations is small, at the cost of less flexible and more biased estimates. More generally, shrinkage estimators introduce bias in a model to improve overall model fit (Efron & Hastie, 2016).

Machine learning research has developed regularization techniques further to cope with complex data structures, such as nonlinear high-dimensional time series data. Latest developments in regularization have extended the principles of the Lasso beyond linear models (Hastie et al., 2016); matrix-based regularization allows to apply regularization techniques to large covariance matrices, for example, time series data or social networks (Tian & Zhang, 2022). Other approaches recover missing or incomplete data when designing recommendation systems (Tian & Zhang, 2022) or integrate regularization in online learning (Farahmand et al., 2008).

Beyond frequentist statistics, the concept of shrinkage also plays a fundamental role in Bayesian estimation. Hierarchical Bayesian estimation pools information from combined observations and thereby shrinks estimates for all observations toward the grand mean (Davis-Stober et al., 2018). Likewise, strong or informative prior distributions for the parameters impose model constraints and thereby serve as a means to regularize the model.

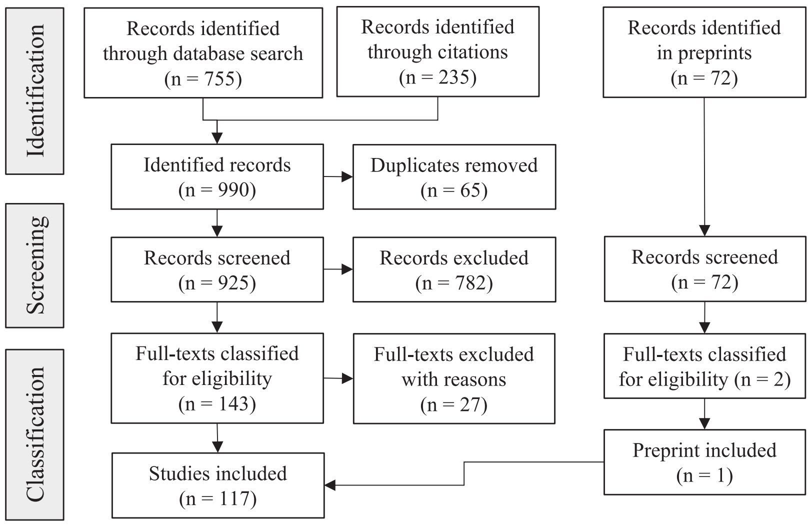
Regularization Within Decision Science: A Scoping Review

I initially aimed to identify in which subfields of decision science regularization methods have left their mark, inspired theory building or advanced applications, and vice versa. To this goal, I conducted a scoping review within the Social Science Citation Index of the Web of Science database covering the time span from 1970 to February 2024. I searched in the title, abstract, and keywords for the search terms regularization OR regularization OR shrinkage OR “lasso regression” OR “ridge regression” OR in combination with the decision science terms (“decision making” OR “decision theory” OR “choice” OR “judgment” OR “judgement” OR “bounded rationality” OR “rationality” OR heuristics). I determined four key publications introducing regularization methods to a wider audience (Girosi et al., 1995; Hoerl & Kennard, 1970; Poggio et al., 1985; Tibshirani, 1996) and searched within their citations for publications in decision science. This initial search identified 755 articles (see Figure 1). Screening of titles and abstracts reduced this set to 143 articles that appeared initially relevant to decision science, dealt with regularization, and were accessible in English. In addition, I searched for openly available preprints published on PsyArXiv via Europe PubMed Central and in the Preprint Citation Index of the Web of Science database within the categories “economics,” “multidisciplinary sciences,” “neurosciences,” “business finance,” “behavioral sciences,” and “psychiatry.” This search identified 72 preprints out of which two preprints fulfilled the stricter eligibility criteria and one preprint was still unpublished. Literature search, exclusion criteria, and the coded database are made available via the Open Science Framework, https://osf.io/dvp37/, (Hoffmann, 2024).

In the classification step, I categorized the articles according to scientific discipline, the methodology used, and the subfield within decision science (see Table 1). If applicable, I determined the theoretical background in decision science (e.g., bounded rationality, prospect theory). I excluded 27 further articles that either did not investigate human decision making (e.g., automated sales forecasts) or did not fall under the broad definition of regularization as adding “some constraints to the minimized objective function, which cannot be obtained from the data and represents the prior preference” (Tian & Zhang, 2022, p. 146).

I further distinguished between three types of articles that differ in the degree to which they seek a theoretical basis for combining decision science with regularization (see Table 2 for a summary of the literature review). Empirical articles use regularization as a tool to better understand human decisions,

Figure 1
Study Identification and Screening



but do not provide a strong theoretical rationale for applying regularization methods. Methodological articles advance the understanding and testing of decision theories with the aid of regularization methods without necessarily establishing a joint theoretical foundation. Vice versa, those articles may advance machine learning methods by integrating concepts from decision science without seeking a

Table 1
Characteristics and Classification of Reviewed Articles N = 117

Characteristic	Article classification
Publication year	Range: 1996–2023; *Mdn* = 2014; *M* = 2013, *SD* = 7.4
Field	Economics and business *N* = 76
	Psychology and social sciences *N* = 18
	Computer science, mathematics, and engineering *N* = 10
	Medicine *N* = 4
	Other *N* = 9
Article type	Review/theoretical integration *N* = 13
	Methodological development *N* = 34
	Regularization as a tool *N* = 70
Area in decision science[a]	Multiattribute decisions and heuristics *N* = 76
	Dynamic decisions *N* = 7
	Forecasting *N* = 29
	Decisions under risk *N* = 19;
	Advice taking and group decisions *N* = 14
	Economic games *N* = 4
	Decision neuroscience *N* = 4
	Axiomatic theories and mathematical modeling *N* = 8
Methodology[a]	Decision experiments with human data *N* = 54
	Simulations *N* = 27
	Secondary data analysis *N* = 40
	Mathematical derivations *N* = 33
Regularization method[a]	l_1-norm (Lasso) *N* = 14
	l_2-norm (Ridge) *N* = 13
	Hierarchical Bayes *N* = 73
	Empirical Bayes/informative priors *N* = 11

[a] Categories are not mutually exclusive.

Table 2

Summary of Previous Approaches to Regularization, Outcomes, and Future Directions

Article type	Empirical (N = 70)	Methodological (N = 34)	Theoretical/review (N = 13)
Rationale	Regularization as a tool	Methodological progress through regularization	Regularization provides theoretical foundation
Theory integration	Low	Low—moderate	High
Approaches	Feature selection; dealing with heterogeneity; solution in economic games	Cognitive modeling of high-dimensional data; mapping descriptive onto normative models; joint modeling of multiple outcomes	Heuristics as priors for full information models; modeling of information constraints; integration of prior beliefs
Outcomes	Interpretability; better generalization; communication	Interpretability; better generalization; improving forecasts of individual judges; adaptive design	Detecting plausible priors for machine learning models; explaining deviations from microeconomic models
Future directions	Application within nonlinear models; stakeholder engagement	Using informative priors in forecasting tools; information constraints in artificial intelligence agents	Integrating decision science with cognitive control; informative priors as means for theory building
Sample article	Barrera Ferro et al. (2020)	Kang et al. (2022)	Galdo et al. (2022)

firm joint ground. Finally, theoretical articles build upon regularization methods to provide a rationale for a theoretical argument within decision science or to further develop a decision theory. In the extraction step, I extracted from all reviewed articles the major insights originating from the integration of decision science with regularization methods and the mentioned directions for future research.[1] I organize the review along the distinction between article types and zoom in on theoretical articles in order to carve out where the concept of regularization has the potential to advance theoretical debates.

Regularization as a Tool in Decision Science

I identified three major approaches to applying regularization as a tool: dealing with heterogeneity, feature selection, and solving economic games. Fifty-five articles implemented hierarchical Bayesian estimation methods to accommodate nested data structures, such as consumers from different households purchasing a set of products from different brands (Arora & Henderson, 2007; Borle et al., 2005). Originating from the need to account for heterogeneous, sometimes extreme preferences of consumers with little available data (Allenby & Ginter, 1995; Lenk et al., 1996), hierarchical Bayesian methods became a standard tool in conjoint modeling of consumer behavior. Applications in decision science beyond marketing, for instance, the modeling of psychological processes, became more prominent after publication of a special issue on hierarchical Bayesian models in cognition (Lee, 2011). These applications resort to hierarchical Bayesian estimation for two reasons. First, it facilitates parameter estimation when it is difficult to derive the likelihood (Terui & Dahana, 2006), when data are missing (Arora, 2006), or only a limited amount of data is available (Bradlow & Rao, 2000). As a result, the estimated parameters are more robust and the model is less prone to overfitting. Second, hierarchical Bayesian estimation allows to efficiently deal with heterogeneity, for instance, varying preferences between households, and to disentangle preferences of individual entities (e.g., households) from preferences on the group level (e.g., specific customer segments or markets). Accounting for the dependencies between data points thereby aids to

[1] Note that I ignore here very specific contributions of the articles to their subfield, for example, which predictors of customer churn were identified within a Lasso regression.

combine and jointly model different information sources (e.g., household debts and assets, Feng et al., 2019), to identify joint preferences (e.g., household dyads, Arora & Allenby, 1999; Arora et al., 2011), or to effectively divide customers into (market) segments (Brand & Baier, 2022; Schreiber, 2017). A limited number of articles applied (empirically) informed priors to increase robustness of the estimation (Terui & Dahana, 2006) or to adaptively select the choice options (Kräplin et al., 2020).

A further set of articles applied regularization methods to select the most important features or predictors for a decision problem. Domains ranged from predicting consumer behavior (e.g., churn De Bock & De Caigny, 2021; Kushwaha et al., 2021; Šimović et al., 2023) to clinical diagnoses and health behavior (Barrera Ferro et al., 2020; Breathett et al., 2019; Hartmann et al., 2009) to plea discounts (Petersen et al., 2022) and movement analysis (Guo et al., 2022). Within these articles, regularization predominantly serves as a tool to prevent overfitting and enhance interpretability. Often, Lasso or Ridge regression are employed as a competitor to nonpenalized regression (Guo et al., 2022; Šimović et al., 2023) or as an initial step to select the variables to be considered in linear and nonlinear modeling (Barrera Ferro et al., 2020; Breathett et al., 2019; Kushwaha et al., 2021). Barrera Ferro et al., for instance, utilized machine learning techniques to understand why attendance levels to public health services are low. After determining the most influential linear predictors, such as elapsed time between the appointment date and the home visit, via a logistic Lasso regression, the reduced set of predictors informed nonlinear modeling with random forest and neural network models. In a last step, the impact of each predictor was communicated via a heatmap to managers of health care facilities in Barrera Ferro et al.'s work. An alternative approach to this two-stage procedure directly integrates the feature selection mechanism into more complex modeling approaches such as spline-rule ensembles (De Bock & De Caigny, 2021). The remaining four articles studied solutions to economic games, such as conflict resolution (Pang et al., 2017), stable cooperative agreements (Gromova & Plekhanova, 2019), or security dilemmas (Albarran & Clempner, 2019). Regularization techniques were implemented to achieve convergence to an optimal solution or to find a stable, time-consistent solution.

Regularization as a Means to Advance Methodological Developments

Methodological articles predominantly applied regularization with the goal to introduce a new methodology to decision science, to improve on existing methods, or to provide a mathematical integration between decision theories. Many articles within this section were motivated from a theoretical perspective within decision science: heuristics and bounded rationality (Davis-Stober et al., 2010; Fokkema et al., 2015; Gilbride & Allenby, 2004; Holzworth, 1996), expected utility (J. Liu et al., 2021; Manchanda et al., 1999; Sugden, 2022), drift diffusion modeling (Kang et al., 2022), prospect theory (Scheibehenne & Pachur, 2015; Toubia et al., 2013), or reinforcement learning (Bastani & Bayati, 2020). I identified two research streams within this set of articles: One stream incorporates regularization techniques into computational choice models with the goal to parsimoniously account for high-dimensional data. For instance, Kang et al. (2022) demonstrated that Lasso regularization may aid to detect brain regions that reliably correlate with model parameters in the drift diffusion model, for example, the drift rate. Again, regularization techniques are predominantly integrated into model estimation because they reduce overfitting to high-dimensional data (Bastani & Bayati, 2020; Chandukala et al., 2011; Gilbride & Allenby, 2004; Kang et al., 2022), render the models more interpretable (W. Chang et al., 2020; Kang et al., 2022), ease the communication of the modeling results (Fokkema et al., 2015), or provide more accurate forecasts for a subgroup of cases (Li & Mayer, 2007). In conjoint analysis, hierarchical Bayesian methods advanced the joint modeling of choice and purchase quantity (Arora et al., 1998), allowed to model the joint maximization of utility and variety seeking (van der Lans, 2018), or to disentangle when promotions led to complementary shopping of consumer products compared to coincidental buying (Manchanda et al., 1999). Last, regularization methods may aid to adaptively select questions or choice options in conjoint modeling (Q. Liu et al., 2009), computational modeling (e.g., prospect theory, Toubia et al., 2013), or online recommendations (Danaf et al., 2019).

Another research stream draws upon regularization as a method to map descriptive models of human decision making onto normative decision theories (see Table 3 for a summary of relations

Table 3
Relations Identified Between Key Concepts in Decision Science and Regularization

Decision concept	Regularization	Relation
Heuristics	l_1-norm	Simple search-and-evaluation rules for consideration sets (Hauser et al., 2010)
	Smart l_2-norm	Combining policies of the judge with the environment structure (Holzworth, 1996)
	Design matrix	Mapping heuristics onto estimators for linear regression (Davis-Stober et al., 2010)
	Informative prior	Prior strength defines continuum from heuristics to linear models (Parpart et al., 2018)
Prospect	Hierarchical Bayes	Nonlinear utility curves can result in aggregated linear utility curves (Kim et al., 2007)
Theory		Loss aversion can be captured without the classical parameter (Nilsson et al., 2011)
Attention	l_1-norm	Attentional limit induces focus on a few dimensions (Gabaix, 2014)
		Attentional limits explain phenomena contradicting microeconomic theory (Gabaix, 2014)
	l_2-norm	Limited total amount of attention implies decay of attention (Galdo et al., 2022)
	Informative prior	Informative priors allow stringent tests of optimal attention (Vanpaemel & Lee, 2012)
Prior beliefs	Informative prior	Integrating probabilities with prior beliefs explains probability weighting (Fennell & Baddeley, 2012)
		Ambiguity expressed as multiple prior beliefs (Grant et al., 2021)
Metalearning	Informative prior	Task costs can be understood as regularization toward default policies (Ritz et al., 2022)
	D_{KL}	Information constraints enforce a higher specialization of single agents (Hihn & Braun, 2020)

Note. D_{KL} = Kullback–Leibler divergence.

between key concepts in decision science and regularization). Along these lines, past work has sought to unify and integrate standard linear models with "improper linear models" or judgment policies (Davis-Stober et al., 2010; Holzworth, 1996). In improper linear models, the importance of each feature is not derived from optimization, but determined heuristically, intuitively, or at random (Dawes & Corrigan, 1974). For instance, equal-weighting models assume that each feature possesses the same importance for predicting the criterion. Holzworth introduce smart ridge regression to clinical decision making as a method to combine improper linear models of the judge with the linear model of the environment. Smart ridge regression incorporates prior information into the penalty term so that regression weights shift toward this prior information instead of zero. As a result, the importance weights of the judge are pulled toward the actual weights in the environment. In contrast, Davis-Stober et al. unify and map improper linear models, such as equal weighting or single cue heuristics, onto regular estimators for linear regression coefficients by imposing constraints via a design matrix approach. This approach allows them to derive statistical properties for improper linear models, for instance, upper bounds for the mean-squared error, and compare their performance systematically to linear regression. Sugden (2022) proposed a more general method to transform revealed choice preferences into a ranking of preferences that is consistent with normative models of choice. To derive these ranked preferences, a planner may specify a regularizing function that maps a descriptive choice model of the judge, for instance, prospect theory, onto a normative theory, for instance, expected utility theory. Taken together, these articles set the seed to integrate heuristic decision rules with optimal, actuarial models of decision making via regularization, a trend that is picked up in newer research.

At times, these methodological advances may revise insights on classical phenomena or generate novel theoretical insights. For instance, if heterogeneity in consumer price sensitivity is not accounted for, conjoint models may overestimate sticker-shock effects, the consternation that consumers experience when shelf prices are unexpectedly high (K. Chang et al., 1999). Hierarchical Bayesian modeling has likewise proven useful as a tool in preferential choice to better understand trade-offs between model parameters and discrepancies between assumptions on the individual level and aggregate behavior. Hierarchical modeling via splines, for instance, can explain how nonlinear utility functions on the individual level can give rise to seemingly linear utility curves on the aggregate level (Kim et al., 2007). Relatedly, implementing prospect theory into a hierarchical Bayesian framework revealed that even without the classical loss aversion parameter, cumulative prospect theory can capture choice patterns typically associated with loss aversion (Nilsson et al., 2011). Subsequent work pointed toward the more general problem that well-established cognitive models suffer from problems of parameter identifiability, leading to interdependencies between model parameters (Scheibehenne & Pachur, 2015). In sum, these articles highlight that regularization methods may help to develop a more profound comprehension of the interplay among model parameters and their theoretical ramifications.

Regularization Informing Theoretical Debates

So far, the majority of the reviewed work has rarely drawn upon regularization methods to make a theoretical contribution to decision science. Recent work has more explicitly applied the concept of regularization to initiate or inform a theoretical debate following three distinct routes: by reinterpreting heuristics within more information-savvy models (Bobadilla-Suarez et al., 2022; Parpart et al., 2018), by reflecting upon the role of prior beliefs in decision making (Fennell & Baddeley, 2012; Gilboa et al., 2012; Grant et al., 2021), and by applying regularization to model information processing and attentional constraints (Gabaix, 2014; Galdo et al., 2022; Hihn & Braun, 2020; Ritz et al., 2022). Building upon earlier approaches (Davis-Stober et al., 2010; Holzworth, 1996) that unify heuristic decision rules with more complex full information models, such as linear regression models, the Bayesian modeling framework can accommodate both model classes by varying the strength of the prior, with stronger priors effectively regularizing the model (Parpart et al., 2018). Linear regression models can be represented within a Bayesian model by extremely weak priors, whereas directed tallying, a heuristic summing up all the evidence for each choice option, is reflected in the same Bayesian model by an extremely strong prior. In principle, these strong priors can be advantageous because they may reflect robust inductive biases that correspond to the structure of the environment (Bobadilla-Suarez et al., 2022). The simple majority rule of tallying, for instance, works sufficiently well whenever features are equally important for making a decision. In line with this idea, Bobadilla-Suarez et al. found that embedding heuristics as priors into regularized regression models outperforms the standard ridge regression approach. Yet, reinstating heuristics as priors within Bayesian models may also offer novel theoretical perspectives on heuristic decision making. For instance, less-is-more effects, that is the effect that sometimes possessing less knowledge can be beneficial, have been taken as support for the adaptive rationality of heuristic decision making (Goldstein & Gigerenzer, 2002). Parpart et al. (2018) argued that less-is-more effects only emerge from ill-specified models within the Bayesian framework. Supporting their argument, they demonstrate that Bayesian models that learn from experience outperform variants with extreme priors in forecasting.

The role of prior beliefs in rational decision making has been further discussed theoretically in the literature on choice under uncertainty and ambiguity (Fennell & Baddeley, 2012; Gilboa et al., 2012; Grant et al., 2021), with a focus on the extent to which prior beliefs can be adequately formalized using Bayesian prior distributions. Starting from the consideration that in an uncertain world statements about probabilities entail a degree of uncertainty, Fennell and Baddeley argue that it is rational to integrate these probability statements with any prior experience, as expressed in prior distributions. While empirical priors allow to harness previous knowledge in situations similar to previously encountered ones, uninformative priors ignore previous experiences and thus more easily accommodate truly novel situations. In reality, individuals often face ambiguous situations, not knowing if a novel situation is similar to a previous one or not, thus multiple priors exist (Grant et al., 2021). Fennell and Baddeley demonstrate that combining empirical and uninformative priors using hyper-priors explains in which situations the subjective perception of probability statements deviates from the objective probabilities, resulting in overweighting of small probabilities and underweighting of large probabilities. Grant et al. further expanded on how ambiguity in language and information can be modeled with multiple prior distributions. Yet, the Bayesian approach to modeling prior beliefs as probability distributions has not been left without critique (Gilboa et al., 2012). Gilboa et al. argue that in many situations, individuals may not have sufficient information to choose an appropriate prior, nor do Bayesian approaches specify theoretically how prior beliefs are generated.

Another recent pathway toward integrating ideas from regularization into decision science considers how attentional constraints can be modeled via regularization techniques (Galdo et al., 2022; Ritz et al., 2022; Vanpaemel & Lee, 2012) and which implications these constraints possess for rational decision making (Gabaix, 2014; Hihn & Braun, 2020; Ritz et al., 2022). Galdo et al. (2022) utilized regularization techniques to test different theoretically plausible mechanisms for attentional constraints within a classical category learning model, the generalized context model (GCM; Nosofsky, 1988). The GCM proposes that individuals rely upon the retrieval of past exemplars stored in memory to categorize newly encountered objects into distinct categories (Medin & Schaffer, 1978; Nosofsky, 1984, 1988). Individuals may learn to selectively attend to more important features via error reduction, but the amount of attention distributed remains fixed across trials. Galdo et al. (2022) contrasted this idea with a) the possibility that limited attention forces individuals to focus attention on a few dimensions, embedding a l1-penalty within the GCM, and b) the possibility that the total amount of attention is limited and causes a decay of attention, embedding a l2-penalty. Across four benchmark data sets and a newly developed experiment, the authors find support that limiting attention to a few dimensions, as suggested by l1-penalty, better describes attentional learning, challenging the traditional assumption that individuals allocate the same amount of attention to the categorization task across trials. Vanpaemel and Lee (2012) demonstrated in a case study of the GCM how informative priors can help to rigorously specify and test a theoretical prediction, namely the optimal distribution of attention across feature dimensions in category learning. While their results imply that the optimal-attention hypothesis only holds for one-dimensional tasks, the case study highlights how informative prior distributions aid to formalize psychological theories and build richer, but less complex psychological models (Vanpaemel & Lee, 2012).

On a metalevel, distributing attention across tasks and selecting cognitive control strategies may itself be framed as an ill-posed decision problem (Ritz et al., 2022). Distinct control strategies, such as enhancing goal-relevant information and suppressing irrelevant information, often allow to reach similar goals. The costs associated with exerting cognitive control, such as task effort, may aid selecting more efficient control strategies via regularization (Ritz et al., 2022).

Information constraints posed by limited attention may subsequently impact on more classical economic choice behavior. In a theoretical review, Gabaix (2014) explored how limited attention, as instantiated via regularization, alters predictions of microeconomic theories of consumer behavior. The developed sparsity-based model of bounded rationality accommodates behavioral phenomena contradicting classical microeconomic theories, such as the money illusion, that is the tendency to value one's wealth in nominal terms instead of its purchase power. Relatedly, Hauser et al. (2010) explored how regularization constraints can enforce cognitively simple, noncompensatory decision rules during the search and evaluation stage in consideration set decisions, thereby allowing to explicitly model trade-offs between finding options with maximum utility and search costs. Finally, Hihn and Braun (2020) reintegrated the idea of bounded-rational agents into multiagent artificial intelligence (AI) systems that jointly solve a complex problem. They explore the possibility that information constraints enforce a higher specialization of the single agents. Simulations demonstrate that when learning to solve a single complex problem, the predictive accuracy of the system increases with the specialization of the single agents. When learning to solve multiple problems, the agents specialize on similar tasks. Interestingly, random noise in information processing, as observed in humans and animals, may serve a similar regularizing function, thereby enabling robust decisions under adverse conditions (Findling & Wyart, 2020). Taken together, these articles suggest that modeling information constraints via regularization may help refine notions of the bounded-rational decision maker and advance our knowledge of which phenomena fall within or outside of the information-constrained decision maker.

Future Directions

With the rise of Big Data, making sense of complex, high-dimensional data have gained importance and, in turn, the integration of regularization methods into common statistical techniques and their application has gained traction. Despite the rising interest in regularization within data science, so far decision science has made limited use of its potential. The scoping review detected only 13 articles that theoretically connected regularization to research topics within decision making, leaving room for future work. The most obvious route is to continue and intensify the application of regularization techniques as a tool across different applied domains from medical decision making (Hartmann et al., 2009) to consumer choice (De Bock & De Caigny, 2021). Similar to previous applications, such approaches may utilize regularization techniques to focus on the most important predictors and to prevent overfitting. Incorporating regularization techniques into the analysis of decision policies may prove beneficial because regularization renders scientific results more understandable to users, practitioners, policy makers, or the general public, both within standard regression-based techniques, such as the Lasso or Ridge regression, as well as more advanced ensemble techniques, such as spline-rule ensembles (De Bock & De Caigny, 2021) or tree-based methods (Fokkema et al., 2015). Regularization may aid to feedback insights from the analysis of decision policies to stakeholders, among them medical staff, patients, or policy makers. A key ingredient to facilitate stakeholder involvement in designing interventions may be to visualize the importance of the predictors selected by regularization (Barrera Ferro et al., 2020), opening up a novel pathway toward successful user engagement in decision research.

More innovative applications may utilize regularization techniques to incorporate informative priors into forecasting tools or into the policy analysis of individual judges (Bobadilla-Suarez et al., 2022; Holzworth, 1996; Parpart et al., 2018). Context-dependent, these priors may exploit expert knowledge and environmental cue validities (Holzworth, 1996) or harness the power of heuristics to generate robust inductive biases (Bobadilla-Suarez et al., 2022; Parpart et al., 2018) Following this spirit, cutting-edge approaches within machine learning demonstrate that equipping regression models with regularization toward the equal-weight heuristic can outperform alternative models when feature directions are unknown (Lichtenberg & Şimşek, 2019). This pathway may be extended to a wider range of heuristics, such as incorporating the take-the-best heuristic as a prior, as well as model classes beyond regression. In turn, it may be feasible to exploit regularization methods as a technique to uncover human priors or detect novel heuristics (Parpart et al., 2018).

Methodologically, as Kang et al.'s work (2022) illustrates, regularization may enable decision scientists to apply and test psychological choice models, such as the drift diffusion model or prospect theory, on high-dimensional data. Beyond testing the models on neuropsychological data, this facilitates applying cognitive choice models to data collected in naturalistic settings, like consumer data (Hornsby & Love, 2022). Vice versa, a principled integration of psychological plausible and ecologically rational information constraints into classical AI algorithms, such as mixture-of-expert models, may aid to develop less task-specific AI agents and boost their ability to generalize across tasks and domains (Bobadilla-Suarez et al., 2022; Hihn & Braun, 2020).

Future theoretical contributions may add to recent work exploring how working memory and attentional constraints can be integrated into psychological decision theories (Gabaix, 2014; Galdo et al., 2022; Hihn & Braun, 2020; Vanpaemel & Lee, 2012). One direction may embed distinct types of attentional constraints into prominent decision theories, such as prospect theory (Tversky & Kahneman, 1981) or decision field theory (Busemeyer & Townsend, 1993), and test their predictions for a set of benchmark phenomena that are supposed to arise from limited attention, such as the money illusion (Gabaix, 2014) or cue competition effects (Busemeyer et al., 1993a, 1993b). An open question that these integrative theories need to address is to what degree the imposed constraints can be conceived as plausible cognitive constructs. It is possible that the absolute magnitude of the penalty term correlates with psychological measures of information constraints, such as working memory capacity, whereas the regularization parameter may be linked to uncertainty one is willing to accept, matching Bayesian theories (Speekenbrink, 2022). Beyond explaining benchmark phenomena, it may be worthwhile to further explore the impact of attentional constraints in online learning (Farahmand et al., 2008; Hoffmann et al., 2019). Capacity constraints imply, for instance, an inconsistent, variable weighting of distinct features, matching empirical findings that individuals with a lower cognitive capacity weigh the cues less consistently (Hoffmann et al., 2014). A systematic investigation of online learning may uncover such phenomena that strengthen the ties between decision science and work on working memory and cognitive control (Ritz et al., 2022).

As the review revealed, regularization methods are not confined to modeling attentional constraints in decision making, but can also aid to spell out how individuals represent prior knowledge (Fennell & Baddeley, 2012), deal with ambiguous information (Grant et al., 2021), or a lack of knowledge (Parpart et al., 2018). More generally speaking, incorporating theoretically motivated constraints into well-established decision theories may allow stricter tests of their assumptions, scope, and falsifiability, thereby advancing theory building and generating a more profound understanding of the psychological processes underpinning human decisions (Vanpaemel & Lee, 2012). In recent years, there has been an increasing advocacy for theory-driven approaches (Oberauer & Lewandowsky, 2019; Lee & Vanpaemel, 2018). Leveraging the capacity of regularization methods, such as classical penalty terms, informative priors, or hierarchical modeling, to formalize theoretical assumptions could present a promising pathway for future theory development.

Discussion

Introduced as an approach to ill-posed problems (Tikhonov, 1963), regularization methods are today standard tools within the machine learning and AI literature (Tian & Zhang, 2022). Although human decisions may be conceived as the prototype for an ill-posed problem, regularization methods have left their trace only in selected areas within decision science. Implementing regularization methods as a tool for conjoint analysis or feature selection stands out as the dominant approach, followed by their application to advance methodology in decision science. Following Tikhonov's original motivation, regularization methods are often employed when dealing with high-dimensional problems and scarce data, aiding to turn "ill-posed into well-posed problems." From a theoretical perspective, the scoping review evidenced close connections to the literature on heuristic decision making, information processing constraints, and prior beliefs. Indeed, key steps in information processing, such as selective attention to relevant features and cognitive control, can be re-interpreted as ill-posed problems (Ritz et al., 2022). However, the rationale for integrating regularization methods into decision theories has often focused on cognitive simplicity, instead of conceptualizing cognitive effort or search costs as problem solutions.

The scoping review aimed to uncover how the concept of regularization shaped decision science and I, therefore, limited this scoping review to articles within the Social Science Citation Index. The first article that referred explicitly to regularization was published 30 years after Tikhonov's work (Holzworth, 1996) and 50% of the publications appeared in the last 3 years. It is hence possible that the scoping review missed out on a few relevant articles that referred to regularization techniques using a different term. However, the interest in regularization techniques within machine learning follows a similar temporal pattern with a surge in research starting from 2012 (Tian & Zhang, 2022). These similar patterns suggest that regularization methods gained popularity across different domains with decision science following suit, potentially because of the increasing demand to create generalizable tools for making sense of high-dimensional data. Combining the expertise of decision scientists, health practitioners, or economists with the capacity of regularized machine learning techniques may open novel avenues for aiding the development of practical decision support systems or decision aids, such as checklists (Winters et al., 2009; Zhang et al., 2021).

It is worth mentioning that the scoping review did not reveal a closer theoretical connection to cost–benefit analysis within decision science (Beach & Mitchell, 1978; Payne et al., 1993) or the evolving literature on resource-rational decision making (Bhui et al., 2021; Griffiths et al., 2015; Lieder & Griffiths, 2020). These research directions pronounce that considering more information within the decision-making process evokes a trade-off between gains in accuracy and additional costs in terms of money, time, or effort. Resource rationality proposes that individuals handle these trade-offs adaptively and formulate expectations at every time step about the degree to which the gain of the next action outweighs the added computational costs (Gershman et al., 2015). Integrating concepts from regularization may advance our understanding of how individuals realize these trade-offs. Specifically, previous studies have successfully traced distinct sources of computational costs (Ritz et al., 2022). Revisiting this research from the perspective of regularization methods may shift the focus toward explaining how individuals can assess these computational costs without imposing another level of computational complexity. Still, adding a penalty term also renders optimization more complex and it may be argued that, as a consequence, these added costs should increase the complexity of information processing. To what degree regularization terms adequately characterize satisficing (Simon, 1955) on the algorithmic and psychological level remains an avenue for future research.

Conclusion

Finding simpler solutions to high-dimensional problems by penalizing for complexity has contributed to the success of machine learning. I am optimistic that regularization methods may inspire decision science to take novel routes in a similar manner: from developing robust forecasting tools that more effectively leverage large naturalistic data sets to improving our understanding of information processing.

References

Albarran, S. E., & Clempner, J. B. (2019). A Stackelberg security Markov game based on partial information for strategic decision making against unexpected attacks. *Engineering Applications of Artificial Intelligence*, *81*, 408–419. https://doi.org/10.1016/j.engappai.2019.03.010

Allenby, G. M., & Ginter, J. L. (1995). Using extremes to design products and segment markets. *Journal of Marketing Research*, *32*(4), 392–403. https://doi.org/10.2307/3152175

Arora, N. (2006). Estimating joint preference: A subsampling approach. *International Journal of Research in Marketing*, *23*(4), 409–418. https://doi.org/10.1016/j.ijresmar.2006.09.001

Arora, N., & Allenby, G. M. (1999). Measuring the influence of individual preference structures in group decision making. *Journal of Marketing Research*, *36*(4), 476–487. https://doi.org/10.1177/002224379903600406

Arora, N., Allenby, G. M., & Ginter, J. L. (1998). A hierarchical Bayes model of primary and secondary demand. *Marketing Science*, *17*(1), 29–44. https://doi.org/10.1287/mksc.17.1.29

Arora, N., & Henderson, T. (2007). Embedded premium promotion: Why it works and how to make it more effective. *Marketing Science*, *26*(4), 514–531. https://doi.org/10.1287/mksc.1060.0247

Arora, N., Henderson, T., & Liu, Q. (2011). Noncompensatory dyadic choices. *Marketing Science*, *30*(6), 1028–1047. https://doi.org/10.1287/mksc.1110.0667

Barrera Ferro, D., Brailsford, S., Bravo, C., & Smith, H. (2020). Improving healthcare access management by predicting patient no-show behaviour. *Decision Support Systems*, *138*, Article 113398. https://doi.org/10.1016/j.dss.2020.113398

Bastani, H., & Bayati, M. (2020). Online decision making with high-dimensional covariates. *Operations Research*, *68*(1), 276–294. https://doi.org/10.1287/opre.2019.1902

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, *3*(3), 439–449. https://doi.org/10.5465/amr.1978.4305717

Bejani, M. M., & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review*, *54*(8), 6391–6438. https://doi.org/10.1007/s10462-021-09975-1

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21. https://doi.org/10.1016/j.cobeha.2021.02.015

Bickel, P. J., Li, B., Tsybakov, A. B., van de Geer, S. A., Yu, B., Valdés, T., Rivero, C., Fan, J., & van der Vaart, A. (2006). Regularization in statistics. *Sociedad de Estadistica e Investigacion Operativa Test*, *15*(2), 271–344. https://doi.org/10.1007/BF02607055

Bobadilla-Suarez, S., Jones, M., & Love, B. C. (2022). Robust priors for regularized regression. *Cognitive Psychology*, *132*, Article 101444. https://doi.org/10.1016/j.cogpsych.2021.101444

Borle, S., Boatwright, P., Kadane, J. B., Nunes, J. C., & Shmueli, G. (2005). The effect of product assortment changes on customer retention. *Marketing Science*, *24*(4), 616–622. https://doi.org/10.1287/mksc.1050.0121

Bradlow, E. T., & Rao, R. (2000). Hierarchical Bayes assortment choice model. *Journal of Marketing Research*, *37*(2), 259–268. https://doi.org/10.1509/jmkr.37.2.259.18733

Brand, B. M., & Baier, D. (2022). Measuring country of origin effects in online shopping implicitly: A discrete choice analysis approach. *International Marketing Review*, *39*(4), 955–983. https://doi.org/10.1108/IMR-03-2021-0139

Breathett, K., Yee, E., Pool, N., Hebdon, M., Crist, J. D., Knapp, S., Larsen, A., Solola, S., Luy, L., Herrera-Theut, K., Zabala, L., Stone, J., McEwen, M. M., Calhoun, E., & Sweitzer, N. K. (2019). Does race influence decision making for advanced heart failure therapies? *Journal of the American Heart Association*, *8*(22), Article e013592. https://doi.org/10.1161/JAHA.119.013592

Busemeyer, J. R., Myung, I. J., & McDaniel, M. A. (1993a). Cue competition effects: Empirical tests of adaptive network learning models. *Psychological*

Science, *4*(3), 190–195. https://doi.org/10.1111/j.1467-9280.1993.tb00486.x

Busemeyer, J. R., Myung, I. J., & McDaniel, M. A. (1993b). Cue competition effects: Theoretical implications for adaptive network learning models. *Psychological Science*, *4*(3), 196–202. https://doi.org/10.1111/j.1467-9280.1993.tb00487.x

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. https://doi.org/10.1037/0033-295X.100.3.432

Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N. C., & Sardanelli, F. (2021). AI applications to medical images: From machine learning to deep learning. *Physica Medica*, *83*, 9–24. https://doi.org/10.1016/j.ejmp.2021.02.006

Chandukala, S. R., Dotson, J. P., Brazell, J. D., & Allenby, G. M. (2011). Bayesian analysis of hierarchical effects. *Marketing Science*, *30*(1), 123–133. https://doi.org/10.1287/mksc.1100.0602

Chang, K., Siddarth, S., & Weinberg, C. B. (1999). The impact of heterogeneity in purchase timing and price responsiveness on estimates of sticker shock effects. *Marketing Science*, *18*(2), 178–192. https://doi.org/10.1287/mksc.18.2.178

Chang, W., Kim, S., & Chae, H. (2020). A regularized spatial market segmentation method with Dirichlet process—Gaussian mixture prior. *Spatial Statistics*, *35*, Article 100402. https://doi.org/10.1016/j.spasta.2019.100402

Danaf, M., Becker, F., Song, X., Atasoy, B., & Ben-Akiva, M. (2019). Online discrete choice models: Applications in personalized recommendations. *Decision Support Systems*, *119*, 35–45. https://doi.org/10.1016/j.dss.2019.02.003

Davis-Stober, C. P., Dana, J., & Budescu, D. V. (2010). A constrained linear estimator for multiple regression. *Psychometrika*, *75*(3), 521–541. https://doi.org/10.1007/s11336-010-9162-8

Davis-Stober, C. P., Dana, J., & Rouder, J. N. (2018). Estimation accuracy in the psychological sciences. *PLOS ONE*, *13*(11), Article e0207239. https://doi.org/10.1371/journal.pone.0207239

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*(2), 95–106. https://doi.org/10.1037/h0037613

De Bock, K. W., & De Caigny, A. (2021). Spline-rule ensemble classifiers with structured sparsity regularization for interpretable customer churn modeling. *Decision Support Systems*, *150*, Article 113523. https://doi.org/10.1016/j.dss.2021.113523

Efron, B., & Hastie, T. (2016). Empirical Bayes. *Computer age statistical inference*. Cambridge University Press. https://doi.org/10.1017/CBO9781316576533.007

Farahmand, A.-M., Ghavamzadeh, M., Szepesvari, C., & Mannor, S. (2008). Regularized policy iteration. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems* (pp. 441–448). MIT Press. https://papers.nips.cc/paper_files/paper/2008/hash/aeb3135b436aa55373822c010763dd54-Abstract.html

Feng, X., Lu, B., Song, X., & Ma, S. (2019). Financial literacy and household finances: A Bayesian two-part latent variable modeling approach. *Journal of Empirical Finance*, *51*, 119–137. https://doi.org/10.1016/j.jempfin.2019.02.002

Fennell, J., & Baddeley, R. (2012). Uncertainty plus prior equals rational bias: An intuitive Bayesian probability weighting function. *Psychological Review*, *119*(4), 878–887. https://doi.org/10.1037/a0029346

Findling, C., & Wyart, V. (2020). *Computation noise promotes cognitive resilience to adverse conditions during decision-making*. bioRxiv. https://www.biorxiv.org/content/10.1101/2020.06.10.145300v1.abstract

Fokkema, M., Smits, N., Kelderman, H., & Penninx, B. W. (2015). Connecting clinical and actuarial prediction with rule-based methods. *Psychological Assessment*, *27*(2), 636–644. https://doi.org/10.1037/pas0000072

Friedrich, S., Groll, A., Ickstadt, K., Kneib, T., Pauly, M., Rahnenführer, J., & Friede, T. (2023). Regularization approaches in clinical biostatistics: A review of methods and their applications. *Statistical Methods in Medical Research*, *32*(2), 425–440. https://doi.org/10.1177/09622802221133557

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, *129*(4), 1661–1710. https://doi.org/10.1093/qje/qju024

Galdo, M., Weichart, E. R., Sloutsky, V. M., & Turner, B. M. (2022). The quest for simplicity in human learning: Identifying the constraints on attention. *Cognitive Psychology*, *138*, Article 101508. https://doi.org/10.1016/j.cogpsych.2022.101508

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias variance dilemma. *Neural Computation*, *58*, 1–58. https://doi.org/10.1162/neco.1992.4.1.1

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278. https://doi.org/10.1126/science.aac6076

Gilboa, I., Postlewaite, A., & Schmeidler, D. (2012). Rationality of belief or: Why savage's axioms are neither necessary nor sufficient for rationality. *Synthese*, *187*(1), 11–31. https://doi.org/10.1007/s11229-011-0034-2

Gilbride, T. J., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, *23*(3), 391–406. https://doi.org/10.1287/mksc.1030.0032

Girosi, F., Jones, M., & Poggio, T. (1995). Regularization theory and neural networks architectures. *Neural Computation*, *7*(2), 219–269. https://doi.org/10.1162/neco.1995.7.2.219

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. https://doi.org/10.1037/0033-295X.109.1.75

Grant, S., Guerdjikova, A., & Quiggin, J. (2021). Ambiguity and awareness: A coherent multiple priors model. *The B.E. Journal of Theoretical Economics*, *21*(2), 571–612. https://doi.org/10.1515/bejte-2018-0185

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229. https://doi.org/10.1111/tops.12142

Gromova, E. V., & Plekhanova, T. M. (2019). On the regularization of a cooperative solution in a multistage game with random time horizon. *Discrete Applied Mathematics*, *255*, 40–55. https://doi.org/10.1016/j.dam.2018.08.008

Guo, Y., Liu, X., Wang, X., Zhu, T., & Zhan, W. (2022). Automatic decision-making style recognition method using Kinect technology. *Frontiers in Psychology*, *13*, Article 751914. https://doi.org/10.3389/fpsyg.2022.751914

Hartmann, A., Van Der Kooij, A. J., & Zeeck, A. (2009). Exploring nonlinear relations: Models of clinical decision making by regression with optimal scaling. *Psychotherapy Research*, *19*(4–5), 482–492. https://doi.org/10.1080/10503300902905939

Hastie, T., Tibshirani, R., & Wainwright, M. (2016). *Statistical learning with sparsity*. CRC Press.

Hauser, J. R., Toubia, O., Evgeniou, T., Befurt, R., & Dzyabura, D. (2010). Disjunctions of conjunctions, cognitive simplicity, and consideration sets. *Journal of Marketing Research*, *47*(3), 485–496. https://doi.org/10.1509/jmkr.47.3.485

Hihn, H., & Braun, D. A. (2020). Specialization in hierarchical learning systems: A unified information-theoretic approach for supervised, unsupervised and reinforcement learning. *Neural Processing Letters*, *52*(3), 2319–2352. https://doi.org/10.1007/s11063-020-10351-3

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hoffmann, J. A. (2024). *A scoping review of regularization in decision science*. https://doi.org/10.17605/OSF.IO/DVP37

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. *Journal of Experimental Psychology: General*, *143*(6), 2242–2261. https://doi.org/10.1037/a0037989

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2019). Testing learning mechanisms of rule-based judgment. *Decision*, *6*(4), 305–334. https://doi.org/10.1037/dec0000109

Holzworth, R. J. (1996). Policy capturing with ridge regression. *Organizational Behavior and Human Decision Processes*, *68*(2), 171–179. https://doi.org/10.1006/obhd.1996.0097

Hornsby, A. N., & Love, B. C. (2022). Sequential consumer choice as multi-cued retrieval. *Science Advances*, *8*(8), Article eabl9754. https://doi.org/10.1126/sciadv.abl9754

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 103). Springer. https://doi.org/10.1007/978-1-4614-7138-7

Kang, I., Yi, W., & Turner, B. M. (2022). A regularization method for linking brain and behavior. *Psychological Methods*, *27*(3), 400–425. https://doi.org/10.1037/met0000387

Kim, J. G., Menzefricke, U., & Feinberg, F. M. (2007). Capturing flexible heterogeneous utility curves: A Bayesian spline approach. *Management Science*, *53*(2), 340–354. https://doi.org/10.1287/mnsc.1060.0616

Kräplin, A., Höfler, M., Pooseh, S., Wolff, M., Krönke, K. M., Goschke, T., Bühringer, G., & Smolka, M. N. (2020). Impulsive decision-making predicts the course of substance-related and addictive disorders. *Psychopharmacology*, *237*(9), 2709–2724. https://doi.org/10.1007/s00213-020-05567-z

Kushwaha, A. K., Kumar, P., & Kar, A. K. (2021). What impacts customer experience for B2B enterprises on using AI-enabled chatbots? Insights from Big data analytics. *Industrial Marketing Management*, *98*, 207–221. https://doi.org/10.1016/j.indmarman.2021.08.011

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7. https://doi.org/10.1016/j.jmp.2010.08.013

Lee, M. D., & Vanpaemel, W. (2018). Determining informative priors for cognitive models. *Psychonomic Bulletin and Review*, *25*(1), 114–127. https://doi.org/10.3758/s13423-017-1238-3

Lenk, P. J., DeSarbo, W. S., Green, P. E., & Young, M. R. (1996). Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science*, *15*(2), 173–191. https://doi.org/10.1287/mksc.15.2.173

Li, Y., & Mayer, W. J. (2007). Impact of corrections for dynamic selection bias on forecasts of retention behavior. *Journal of Forecasting*, *26*(8), 571–582. https://doi.org/10.1002/for.1028

Lichtenberg, J. M., & Şimşek, Ö. (2019). Regularization in directable environments with application to tetris. In K. Chaudhuri & R. Salakhutdinov (Eds.), *Proceedings of the 36th international conference on machine learning* (pp. 3953–3962). Proceedings of Machine Learning Research.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, Article e1. https://doi.org/10.1017/S0140525X1900061X

Liu, J., Kadzinski, M., Liao, X., & Mao, X. (2021). Data-driven preference learning methods for value-driven multiple criteria sorting with interacting criteria. *INFORMS Journal on Computing*, *33*(2), 586–606. https://doi.org/10.1287/ijoc.2020.0977

Liu, Q., Dean, A., Bakken, D., & Allenby, G. M. (2009). Studying the level-effect in conjoint analysis: An application of efficient experimental designs for hyper-parameter estimation. *Quantitative Marketing and Economics*, *7*(1), 69–93. https://doi.org/10.1007/s11129-008-9045-9

Manchanda, P., Ansari, A., & Gupta, S. (1999). The “shopping basket”: A model for multicategory purchase incidence decisions. *Marketing Science*, *18*(2), 95–114. https://doi.org/10.1287/mksc.18.2.95

McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 93–100. https://doi.org/10.1002/wics.14

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Moradi, R., Berangi, R., & Minaei, B. (2020). A survey of regularization strategies for deep models. *Artificial Intelligence Review*, *53*(6), 3947–3986. https://doi.org/10.1007/s10462-019-09784-7

Nilsson, H., Rieskamp, J., & Wagenmakers, E. J. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, *55*(1), 84–93. https://doi.org/10.1016/j.jmp.2010.08.006

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104–114. https://doi.org/10.1037/0278-7393.10.1.104

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin and Review*, *26*(5), 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

Pang, J. S., Sen, S., & Shanbhag, U. V. (2017). Two-stage non-cooperative games with risk-averse players. *Mathematical Programming*, *165*(1), 235–290. https://doi.org/10.1007/s10107-017-1148-1

Parpart, P., Jones, M., & Love, B. C. (2018). Heuristics as Bayesian inference under extreme priors. *Cognitive Psychology*, *102*, 127–144. https://doi.org/10.1016/j.cogpsych.2017.11.006

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

Petersen, K., Redlich, A. D., & Wilson, D. B. (2022). Where is the evidence? Comparing the effects of evidence strength and demographic characteristics on plea discounts, *Journal of Quantitative Criminology*, *39*, 919–949. https://doi.org/10.1007/s10940-022-09555-8

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, *317*, 314–319. https://doi.org/10.1038/317314a0

Ritz, H., Leng, X., & Shenhav, A. (2022). Cognitive control as a multivariate optimization problem. *Journal of Cognitive Neuroscience*, *34*(4), 569–591. https://doi.org/10.1162/jocn_a_01822

Santos, C. F. G. D., & Papa, J. P. (2022). Avoiding overfitting: A survey on regularization methods for convolutional neural networks. *ACM Computing Surveys*, *54*(Suppl. 10), Article 213. https://doi.org/10.1145/3510413

Scheibehenne, B., & Pachur, T. (2015). Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin and Review*, *22*(2), 391–407. https://doi.org/10.3758/s13423-014-0684-4

Schreiber, F. (2017). Identification of customer groups in the German term life market: A benefit segmentation. *Annals of Operations Research*, *254*(1–2), 365–399. https://doi.org/10.1007/s10479-017-2446-y

Simon, H. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

Šimović, P. P., Chen, C. Y. T., & Sun, E. W. (2023). Classifying the variety of customers’ online engagement for churn prediction with a mixed-penalty logistic regression. *Computational Economics*, *61*(1), 451–485. https://doi.org/10.1007/s10614-022-10275-1

Speekenbrink, M. (2022). Chasing unknown bandits: Uncertainty guidance in learning and decision making. *Current Directions in Psychological Science*, *31*(5), 419–427. https://doi.org/10.1177/09637214221105051

Sugden, R. (2022). Debiasing or regularisation? Two interpretations of the concept of ‘true preference’ in behavioural economics. *Theory and Decision*, *92*(3–4), 765–784. https://doi.org/10.1007/s11238-022-09876-x

Terui, N., & Dahana, W. D. (2006). Estimating heterogeneous price thresholds. *Marketing Science*, *25*(4), 384–391. https://doi.org/10.1287/mksc.1050.0181

Tian, Y., & Zhang, Y. (2022). A comprehensive survey on regularization strategies in machine learning. *Information Fusion*, *80*, 146–166. https://doi.org/10.1016/j.inffus.2021.11.005

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tikhonov, A. N. (1963). On the regularization of ill-posed problems. *Doklady Akademii Nauk SSSR*, *153*(1), 49–52.

Toubia, O., Evgeniou, T., Johnson, E., & Delquié, P. (2013). Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science*, *59*(3), 613–640. https://doi.org/10.1287/mnsc.1120.1570

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

van der Lans, R. (2018). A simultaneous model of multiple-discrete choices of variety and quantity. *International Journal of Research in Marketing*, *35*(2), 242–257. https://doi.org/10.1016/j.ijresmar.2017.12.007

Vanpaemel, W., & Lee, M. D. (2012). Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin and Review*, *19*(6), 1047–1056. https://doi.org/10.3758/s13423-012-0300-4

Winters, B. D., Gurses, A. P., Lehmann, H., Sexton, J. B., Rampersad, C., & Pronovost, P. J. (2009). Clinical review: Checklists—Translating evidence into practice. *Critical Care*, *13*(6), Article 210. https://doi.org/10.1186/cc7792

Zhang, H., Morris, Q., Ustun, B., & Ghassemi, M. (2021). Learning optimal predictive checklists. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. Wortman Vaughan (Eds.), *Advances in neural information processing systems* (pp. 1215–1229). Curran Associates.

Received May 30, 2023
Revision received April 3, 2024
Accepted May 20, 2024 ■