Journal for Nature Conservation 74 (2023) 126447

Contents lists available at ScienceDirect

Journal for Nature Conservation

journal homepage: www.elsevier.com/locate/jnc

Aggregated mental models predict observed outcomes following Eurasian Beaver (*Castor fiber*) reintroduction

Andrew Blewett [*], Maarten Jacobs, Kasper Kok

Wageningen University & Research, Netherlands

ARTICLE INFO

Keywords:
Mental models
Prediction
Conservation
Reintroductions
Wildlife
Castor fiber

ABSTRACT

Outcome prediction is important for conservation; however, analysis may be hampered by specialist resource deficiencies. Mental modelling techniques offer a potential solution, drawing on accessible sources of knowledge held informally by local stakeholders. Mental models show linked social and ecological variables from the perspectives of community members, whose insights may otherwise be neglected. Currently, an important weakness in conservation mental modelling is inadequate attention paid to real-time model predictive validity. To address this knowledge gap, baseline mental model predictions concerning Beaver (*Castor fiber*) reintroduction in Southwest England were followed up at three years. Participants were invited to submit outcome observations for concept variables identified in their original models, blind to inferences based on model dynamic analysis, so that the two sets of data could be compared. Individual concept values and models were found to show weak and highly inconsistent predictive validity, however, multi-stakeholder aggregated mental models showed consistently strong predictive performance. This finding was enhanced by setting tighter thresholds for inclusion of individual model items in aggregation procedures. Threshold effects can be interpreted as a reflection of greater agreement: tighter thresholds retain more highly shared model components. It is proposed that enhanced real-time predictive validity for aggregated models is explained by a 'wisdom of the crowd' statistical effect, analogous to well-recognised crowd judgement effects observed in relation to much simpler questions. The findings show the scope for stakeholder mental modelling methods as an investigative tool, to supplement more conventional ecosystem assessments in predicting data-poor conservation outcomes.

1. Introduction

1.1. Conservation background to the study

Conservation planners and managers are often expected to forecast the results of interventions including reintroductions, wildlife protection measures and ecological restoration, increasingly typical of the urgent large-scale vision of recovery required to counteract ecosystem decline. To assist with the technical and resource challenges involved, stakeholder perception-based mental model methods have been developed as a complement and sometimes as an alternative to conventional ecosystem assessments (Biggs et al., 2008; Jones et al., 2011; Moon et al., 2019). As such, mental models drawing on community knowledge can be elicited to show conceptual components and relationships, usually represented on two-dimensional maps, analogous to their adaptive function in normal psychology (Johnson-Laird, 2010). Represented externally, mental models reveal insights into perceived ecosystem states, and thus help people make sense of their own and others' perspectives and expectations (Jones et al., 2011). How well mental model predictions are borne out by subsequent observations remains untested. This paper aims to help bridge this knowledge gap by examining the question empirically.

A defining feature of mental modelling is capacity to make dynamic "predictions of future system states" (Rouse & Morris, 1986, p351). Predictive function is key to their purpose. As originally conceived, mental models support decision-making by allowing the subject to envisage and select between desirable and undesirable future predictive options (Craik, 1943). In perception-based research, mental models can be formatted as semi-quantitative 'fuzzy cognitive maps' (FCM) to make dynamic predictions which similarly set out alternative predictive future scenarios (Jetter & Kok, 2014). The potential of the technique has been shown in studies of policy in threatened Brazilian forest (Kok, 2009),

* Corresponding author at: Wageningen University & Research, Cultural Geography Chair Group, Wageningen Campus, Building 101 (Gaia), Droevendaalsesteeg 3, 6708PB Wageningen, Netherlands.
E-mail address: andrewearlston.blewett@wur.nl (A. Blewett).

https://doi.org/10.1016/j.jnc.2023.126447
Received 13 November 2022; Received in revised form 3 June 2023; Accepted 29 June 2023
Available online 1 July 2023

and comparing then combining perspectives on the Atlantic Summer Flounder, *Paralichthys dentatus* fishery (Gray et al., 2012).

In mental model analysis, important differences exist between insights gained from individual and aggregated models. Both individual models and combined models from a single set of similar stakeholders should be considered "incomplete representations of 'reality' that are context dependent …" (Moon et al., 2019, p2), primarily of use to shed light on a distinctive point of view within a wider system. In contrast, multi-stakeholder aggregations drawing on more diverse sources of information have been shown to tend towards an objectively verifiable reality for the system under examination (Gray et al., 2012).

The use of combined data to improve the accuracy of real-world assessments has a long history. Galton (1907) first proposed combined judgements based on multiple observations for a simple task such as guessing the weight of a body, in which average results out-performed individuals including recognised 'experts'. More complex predictive cognitive problems implicating diverse fields including economics, politics, sport and engineering for which multiple inputting factors are relevant, have also been shown amenable to crowd estimation effects (Surowiecki, 2005). Elaborating on Surowiecki's theoretical insights, Estupiñán Ricardo et al. (2020) demonstrated a detailed mathematical approach to calculating uncertain relationships in a crowd-sourced FCM, based on averaged crowd estimations for each connection value.

Termed 'wisdom of the crowds' (WOC), multiple source averaging methods have been adapted to complex problem evaluation in conservation (Arlinghaus & Krause, 2013). Examples show the potential for WOC methods applied to challenging wildlife management problems; for example, assessments of the declining abundance of culturally valued Manus Green Tree Snails, *Papustyla pulcherrima* in Papua New Guinea (Whitmore, 2016), and abundance of two threatened species of Asian Horse-shoe Crab, *Tachypleus tridentatus* and *Carcinoscorpius rotundicauda* off the coast of Guanxi Zhuang in southern China (Liao et al., 2019).

Building on this approach, a WOC effect drawn from combined mental models has recently been demonstrated for a European freshwater fishery, where aggregated mental models were found to perform well against independent assessments (Aminpour et al., 2020). To our knowledge, no attempt has been made to take the next step in reflecting back on the psychological forward-predictive role of mental models, applying WOC statistical inference to real-time predictive dynamic analysis of aggregated models in conservation.

The present study compares mental model dynamic analysis with participant follow-up evaluations of the actual 'outcome' activity state for concept components identified in the original models. By this is meant whether a concept variable becomes more or less prominent in the participant's evaluation of the ecosystem. We anticipated that a 'crowd' combination of diversely sourced models would replicate WOC 'averaging' effects, hence enhancing predictive validity for concept activation levels of aggregated models under dynamic analysis.

Problematically for this method, defining a predictive 'average' mental model is not straightforward. A possible solution is to consider a rising set of model concept and connection frequency-inclusion thresholds in aggregation. The effect is to retain only those components showing increasing commonality and hence a plausible expression of 'average' content in a combined model. In practice, it is important to understand how aggregation procedures affect the predictive validity of dynamic analysis, a question which can be explored by comparing model predictive validity while progressively tightening aggregation inclusion criteria.

1.2. The River Otter Beaver Trial

The opportunity to study mental model predictive validity followed earlier work on mental models and the reintroduction of a population of free-living Eurasian Beaver, *Castor fiber* into southwest England, the River Otter Beaver Trial or ROBT, (Brazier et al., 2020). The social dimension of ROBT has been extensively explored including approaches to human-wildlife conflict (Auster et al., 2021), perceived effectiveness of conservation mitigation actions (Blewett et al., 2021), and the role of emotion in responding to a changing ecosystem (Blewett et al., 2022). Beaver ecology is a rich focus for social dimensions of conservation research, as the beaver keystone function triggers rapid ecological change, impacting strongly on landscapes typical of present and former natural beaver range in northern Europe (Brazier et al., 2021). Given these effects, restoring beavers to England has unsurprisingly stimulated vigorous public debate, including how best to promote successful coexistence (Auster et al., 2020), highlighting categories of frequently identified human wildlife conflict applicable to a wide range of global conservation scenarios (Nyhus, 2016).

1.3. Mental model aggregation and dynamic analysis

Construction of mental models is most commonly done by participant interviews followed by pre-specified rules for aggregation (Gray et al., 2014; Özesmi & Özesmi, 2004). Of note, as an alternative to aggregation, it is possible to build multi-stakeholder environmental FCM models using a participatory workshop format such as the method described by Verkerk et al. (2017). As the original baseline study focussed on individual mental models collected primarily to understand individual and sectoral perspectives (*e.g.*, farmers compared with conservationists), this methodology was not a later option to consider with the available dataset. The separate question of the predictive validity of appropriately constituted workshop FCMs is therefore outside the scope of this paper. We formatted FCMs from raw mental models using so-called 'fuzzy logic' methods to cope with subjectivity in judging causal influences between concepts (Kosko, 1986). Connection values from multiple models in FCM format can be combined mathematically. Dynamic analysis (DA) is the predictive procedure whereby the collective strengths of connections are repeatedly applied to the 'state values' of the concepts in the model, which consequently evolve towards a steady state profile representing their relative activity in the system as perceived by the modeller. This can readily be done using one of several freely available on-line software products.

1.4. Mental model predictions and the problem of time

Informal knowledge-based mental models may include beliefs about time scales, however neither mental models nor FCMs standardise time. Punctual concept changes such as adjustments to the law are handled in the same way as 'slow' concepts such as climatic warming. Consequently, semi-quantification of both state and temporal evolution of the system requires care in interpreting meaning of change, because dynamic analysis of FCMs is entirely relative, *e.g.*, concepts are more or less present and or active, *post* interaction with the rest of the model. This non-specificity challenges the notion of model prediction as a future forecast. An additional challenge arises from accumulating events. For example, the present study follow-up period included both the traumatic rupture in U.K. relationships with the European Union ('Brexit'), and imposition of Covid pandemic restrictions to social and economic activity. Elicited mental models have no in-built capacity to adjust for such novelties, and so one might expect 'predictive decay' as a result of incidental events over time. A prospective follow-up study embracing major disruptions over such a short time frame offers a unique opportunity to assess mental model predictive robustness in the face of system shocks.

1.5. Research hypotheses

This study investigates the role of mental modelling in evaluating expected change within a conservation landscape. Although the species focus is on Eurasian beavers, the method has wider applications in conservation science and practice. To our knowledge future predictive mental modelling is as yet unaddressed in this context (Conservation evidence base, https://www.conservationevidence.com).

Thus, applying the potential value of aggregated mental modelling as a WOC tool to our case study of a beaver reintroduction in a typical multi-use lowland north European landscape, we proposed hypotheses H1 and H2;

H1. FCM dynamic analysis of baseline stakeholder conservation mental models predicts subsequent follow-up observations, and

H2. Aggregated FCM mental model predictive potential is superior to equivalent tests for either individual mental models or single concepts composing individual models, consistent with 'wisdom of the crowd' logic.

Additionally, to address the methodological issue of how stringency of aggregation techniques influences the predictive validity of combined models, we pose a research question, RQ;

RQ. How do decisions concerning inclusion thresholds for concepts and connections by degree of 'sharedness' between participants influence the predictive potential of multi-stakeholder aggregated models?

2. Method

2.1. Study sample

This study builds on a baseline sample of 48 participant mental models recruited over six months by the principal author in 2018/19 (see Blewett et al., 2021), of which 31 provided usable follow-up data three years later.

The study considered the ROBT geographical area. The river Otter runs south into the western English Channel, from a small catchment of approximately 250 Km^2 on the Southwestern English peninsula. At the end of the ROBT in 2020, resident beavers formed a growing population approaching 50 animals within the catchment. The river passes through a mixed lowland and mostly agricultural region; riparian areas comprising grassland used for grazing, smaller areas of arable, woodland, and small village and urban settlements with connecting infrastructure.

Study participants were recruited on the basis of active engagement with the study area, and initially interviewed in their homes or workplace. Stakeholder affiliations were identified by discussion with the ROBT project manager, and through annual reports later summarized by Brazier et al. (2020). Categories include General Public, Conservation and Environmental Scientists, Landowners and Managers, Farmers, Anglers and a miscellaneous group of regulators and natural resource managers. 23 people, including most specialist participants were recruited by direct approach or snowballing, and 25 including most general public participants via social media adverts. As the focus is on knowledge, recruitment aimed "not to obtain a representative sample of a population, but to represent different knowledge areas", (Olazabal et al., 2018: 800) and hence sampling up to 'concept saturation' (Özesmi & Özesmi, 2004) which arrived at 89 % by 25 interviews; however, recruitment was continued to better represent minority specialist stakeholder groups.

Mental model interviews included an introduction to the study, with sharing photographs of a beaver, the river and trial area map. In addition to mandatory 'Beaver presence', participants were asked to write down and display their own 'concepts' framed under three categories, while

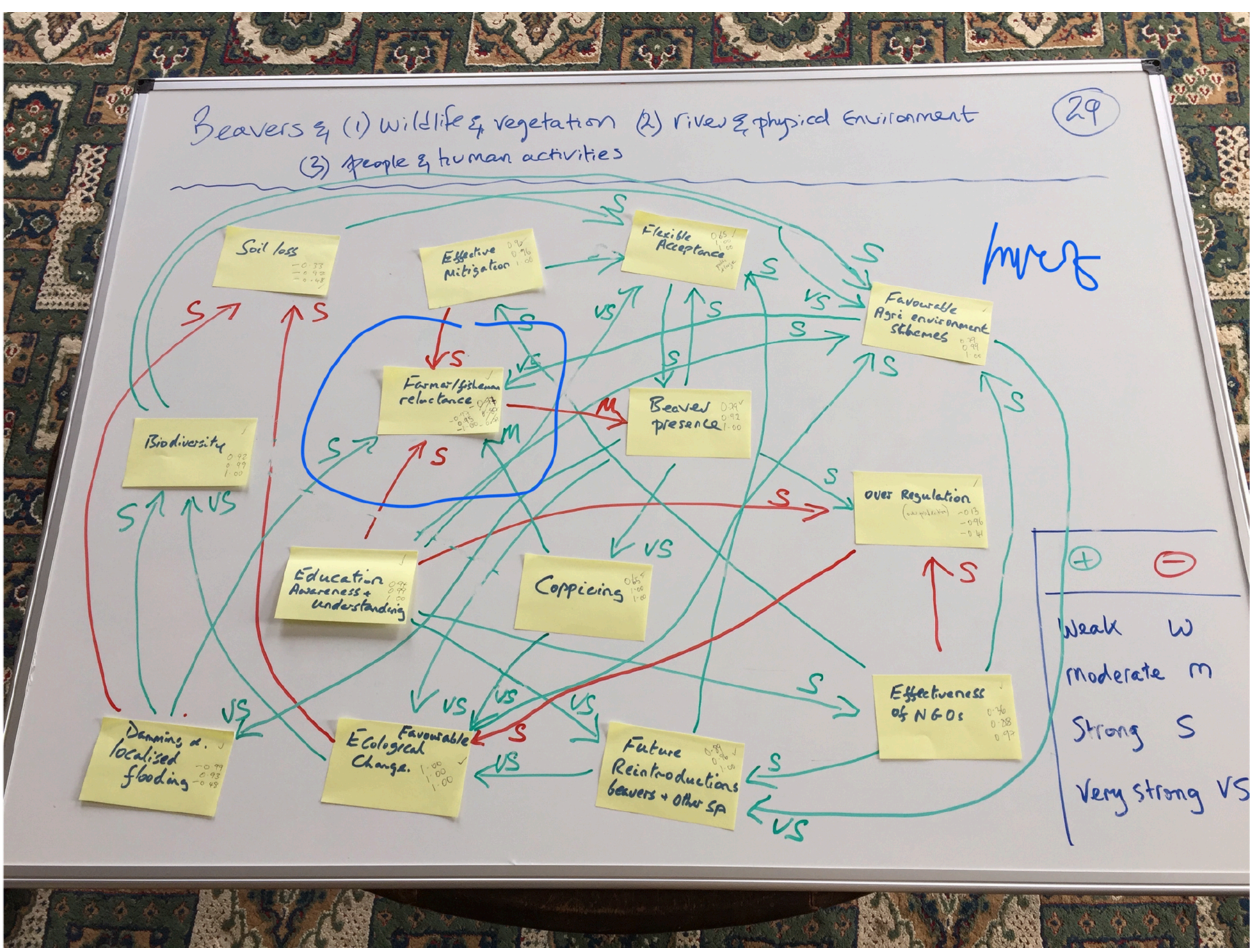

Fig. 1. Example of a (Landowner) raw mental model. One concept was inverted for semantic consistency, its connections inverted in the FCM matrix for analysis to preserve the logic of the model. With kind permission.

considering a five-year time span: 'Beavers and a) wildlife & vegetation, b) river & physical environment c) people & human activities', using positive terms wherever possible. When the concept display was complete, participants were asked to add direct (+) or inverse (−) fuzzy-weighted connecting 'influence' arrows (Very Strong 'VS', Strong 'S', Moderate 'M', and Weak 'W') starting from each concept in turn. The objective was to obtain representations as close as possible to the working mental model held by each participant, for the relevant domain; see for example, Fig. 1.

2.2. Mental model representation as FCMs

Initial cleansing reduced the raw sample from 657 to 600 mental model terms, after replicates on the same models were merged. 22 individual terms plus 2 merged terms were inverted to express a positive meaning to facilitate comparison and aggregation.

Mental model representation for FCM analysis follows published studies (Obiedat & Samarasinghe, 2016; Olazabal et al., 2018; Özesmi & Özesmi, 2004): (i) 'raw' concepts were standardised as 53 'condensed' FCM concept-categories based on semantic similarity, (ii) FCMs were converted into square matrices, cells populated with numerical values equating to 'fuzzy' terms, (no influence = 0, W = ±0.2, M = ±0.4, S = ±0.6, VS = ±0.8), and for aggregated models, (iii) mean connection values were calculated for concatenated cell concept connections (Abel et al., 1998; Cannon-Bowers & Salas, 2001; Gray et al., 2014), following which, (iv) inferred concept activation states influenced by connections were calculated by dynamic analysis (DA).

2.3. Dynamic analysis, DA

Baseline DA was conducted using open access 'FCMapper' software (Wildenberg et al., 2010) for (i) 31 individual mental models for which participants agreed to follow-up, and (ii) ten versions of the 31-model aggregated FCM, determined by setting two concept inclusion thresholds and five connection inclusion thresholds.

DA is done by allowing the vector set of 'activation values' for each concept in the model (A_1, A_2, A_2..., A_n) to evolve under iterative multiplication with the constant matrix ($n \times n$) of signed and weighted connections, until the vector set values stabilise. An initial standard value of '1' is granted to all concepts comprising vector A at time-point k, the connection matrix w is composed of rows i and columns j. Multiplication arrives at $k + 1$, repeated until the values stabilise. No 'self-loops' ($i = j$) were permitted. The procedure is set out in Equation (1);

$$A_i^{(k+1)} = \Sigma^N_{\substack{j \neq i \\ j = 1}} A_j^{(k)} w_{ji} \qquad (1)$$

FCMapper includes the option of logistic normalisation for each concept activation step. We analysed non-normalised data for aggregated models, to avoid an unnecessary additional step in data transformation.

2.4. Follow-up observed data

Follow-up comprised 'Google Forms' questionnaires following favorable email response to a request for participation covering views on how each concept identified in the original models had fared subsequent to baseline mental model interviews, as a point observation at approximately three years (range: 36–43 months). 32 forms were eventually completed, of which 31 could be analysed.

Occupational, gender and age distributions are shown in Table1.

Of note, three conservationists worked directly with the ROBT and two members of the general public had family links, however these individuals were judged to be working and thinking independently, with different professional and personal priorities expressed through their mental models. No attempt was made to compare predictive validity by stakeholder group, as the aggregation focus was on the whole sample; the most diverse representation of stakeholder knowledge relevant to the principles of the WOC statistical effect.

Overall gender balance, thirteen women and eighteen men, showed marginal male over-representation, probably reflecting community gender bias in farm ownership and forestry, and angling as a leisure activity, contrasting with a more balanced younger female age-group involvement in conservation, supported anecdotally. Female predominance amongst interested members of the public is compatible with the possibility that women are poorly represented in some rural operational areas of control (*e.g.*, of land use) but show a strong level of concern from 'outside'. The demographics are likely to be a simple reflection of gender-biased social roles in the study region. Ethnicity data was not collected. It is unlikely that any participants would have identified as other than white British or European.

The follow-up questionnaire comprised a 7-point (-3 to 3) scale for perceived change to each term, according to one of seven labels; "Decreased a lot", "Decreased moderately", "Decreased a bit", "Unchanged", and conversely step-wise to "Increased a lot", with an eighth "Don't know" option where preferred. Participants did not have DA results available either from their own or the aggregated model, hence participant estimates were 'blind' to baseline DA, and dependent only on their updated perspectives on the developing beaver ecosystem.

To complete the questionnaire, participants were instructed to consider change to concept activity, extent, and/or quality as applicable and irrespective of reason, in the three years since 2018/19. They were asked to use broad judgement to arrive at a net summary view reflecting experience, information or knowledge; stressing that the focus was on perception and not a supposed right or wrong response. Participants were asked to grade outcomes for their original mental model terms, reverted back to the standardised concept label for aggregation.

2.5. Model predictive potential

Predictive potential of the aggregated models was tested by regression analysis for the set of predictive DA concept activation values and the corresponding set of mean follow-up questionnaire scores at three years. Activation values for concepts with outgoing connections only (transmitters) were excluded from correlation analysis because as static drivers, they cannot themselves show change in a future state of the system.

Individual concept predictions and observations both show non-normal distributions on respective scales of concept activity change. Aggregated models generate relatively small data sets. All follow-up outcome observation data are measured on an ordinal scale. For these reasons non-parametric Spearman's ρ was considered most appropriate

Table 1
Demographic structure of the follow-up sample, n = 31.

	General public	Conservation and environmental science	Landowners and estate managers	Farmer and farmer representative	Angler	Forestry managers and regulators
Number of mental models	14	8	3	2	2	2
F:M gender	9:5	4:4	0:3	0:2	0:2	0:2
Age range	26–61+	18–60	41–60	41–61+	18–61+	26–60

to test all predictive concept activation and outcome score correlations, and ρ^2 to estimate proportions of explained variance.

2.6. Non-aggregated data

All 31 follow up mental models were dynamically analysed in FCMapper, independently of each other, such that individual model concept DA, and the whole sample of concept DA scores, could be compared with corresponding follow-up observations. In analysis of the whole sample of concept DAs, concept DA values were normalised because outcome values can otherwise only be compared within models but not between models.

2.7. Aggregation: Threshold setting procedure for concepts

There is no standardised approach to identifying a common set of concepts retained in aggregation. Aggregation methods include retaining all concepts and stronger summated connections (Özesmi & Özesmi, 2004), and inclusion by expert-credibility weighting (Obiedat & Samarasinghe, 2013). We relied on frequency of concept inclusion by contributing stakeholders (see Blewett et al., 2022), retaining the most commonly identified concepts above the following thresholds; (i) a 'permissive' cut-off retaining 25 concepts present in the upper two quartiles by concept frequency and excluding a long tail of rarer outlier concepts, and (ii) a 'tight' cut-off retaining 12 concepts present in the upper quartile only, to favour a much higher degree of concept-sharedness.

The terms 'permissive' and 'tight' are used to distinguish concept inclusion thresholds hereafter. As connection thresholds are themselves tightened, rarer, less-shared concepts lose connections and are progressively orphaned, falling out of the aggregated model.

2.8. Aggregation: Threshold setting procedure for connections

Mean connection values between retained concepts were calculated in Microsoft Excel for permissive and tight concept versions of the aggregated model. The procedure retains all same-signed (meaning positive-direct or negative-inverse causality) connections between equivalent concepts. The threshold selections chosen require ≥2, 3, 4, 5 and 6 connections in the relevant concatenated FCM matrix 'cell', from which a final mean value is calculated. It was planned that each group would be extended to include a higher minimum, accepting no more than one contrary signed member which we anticipated we would then drop as an outlier in calculating a mean value; however, no such cases arose.

3. Results

3.1. H1, individual concepts and models

Predicted change (using normalised DA values) was associated with observed change ($\rho = 0.25$, $P < 0.001$) for the whole sample of concepts ($N = 361$ pairs), conventionally considered to be a weak effect at $0.3 \geq \rho > 0.1$ (Cohen, 1992). Overall, only 6 % of the variance in observed change was explained by the predicted change (*i.e.*, $\rho^2 = 0.06$).

Individual models analysed by non-normalised DA showed a very wide predictive spread. For six of 31 individual models, predicted change was strongly correlated with observed change ($\rho \geq 0.50$, $P < 0.05$). For six other individual models, the association was negative. The six individuals with strong positive correlations were from four different stakeholder groups. Overall, individual model ρ values lie in the range −0.41 to +0.77.

3.2. H2, the aggregated model, with example

Relationships between predicted changes (DA) and observed changes were tested for aggregated models based on two concept and five connection thresholds, generating ten versions of the aggregated model in total.

The 'tightest', most strongly predictive aggregated mental model, (version 10, see Tables 2 and 4) has the highest thresholds for inclusion based on concepts shared or most frequently found amongst the thirty-one contributing individual mental models. Consequently, it has twelve concepts, of which three have lost all inputting connections because of the high connection threshold. The state value of these driver concepts is static, hence excluded from correlation calculations. For illustration, in

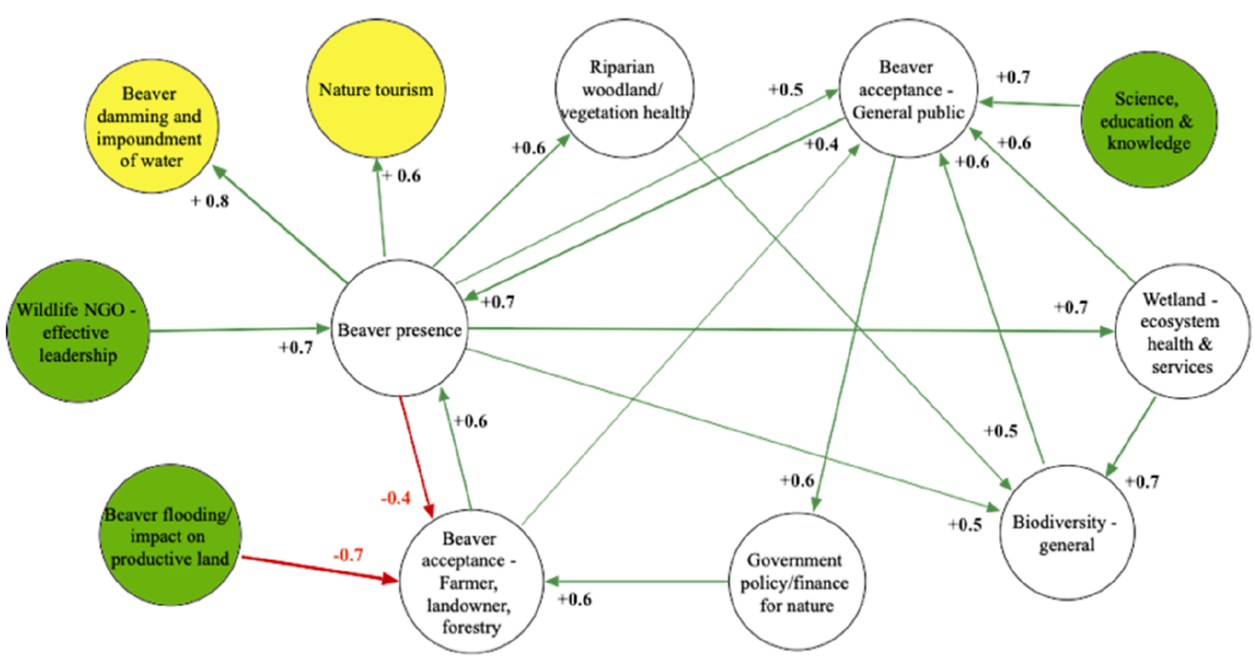

Fig. 2. Illustrative aggregated mental model; version 10, shown in Tables 2 and 4. Transmitter (driver) concepts are green, receiver concepts yellow, concepts with both in and outputs are white. Connection values are direct - black, or inverse - red; all connection values (−1 to +1), are rounded to one decimal place for clarity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

addition to 'Beaver presence', model 10 shows the participant originated concepts 'Beaver acceptance – General public', 'Beaver acceptance – Farmer, landowner, forestry', and 'Biodiversity – general' as most connected (more incoming and outgoing influence arrows) and hence considered most salient in relation to the whole model; see Fig. 2.

The relationship between predictive dynamic analysis concept state values in model version 10, are shown in relation to observed values representing mean participant perceptions of change in the state of the concepts at three years, in Table 2.

Equivalent findings for all ten versions of the aggregated model examined are shown for permissive and tightly defined concept inclusion aggregated models in Tables 3 and 4 respectively. Each version shows strong predictive effects, considered as $\rho \geq 0.5$ (Cohen, 1992), significant at $P < 0.05$ in all but model version 4. Across all aggregated models, between 38 % and 72 % of variance in observed change is explained by the predicted change (*i.e.*, $\rho^2 =$ between 0.38 and 0.72).

Predictive performance measured by ρ tends to increase as concept inclusion is tightened, demonstrated by the spread of ρ values compared between the permissive and tight concept inclusion models shown in Table 3 and 4. There is also a rising pattern of ρ values corresponding to increasing connection thresholds shown in the tight concept inclusion model, Table 4. Spearman's ρ calculations excluded concepts which function as transmitters, i.e., only have outgoing connections and hence cannot be influenced within the model, which appear in aggregated model versions 9 and 10, also shown in Table 4.

3.3. Interaction between connection and concept inclusion thresholds in aggregation

Aggregated model DA values shown in rows for individual concepts and the corresponding mean observation value at three years are set out in Tables 5 and 6, again distinguishing between permissive and tighter thresholds based on concept frequency across the sample of individual mental models. The concept 'Beaver presence' was mandatory for all participants at elicitation, and so can be seen to occur in all 31 individual mental models. Columns of DA values for increasing connection thresholds are correlated with mean observations, (from which are derived the Spearman's ρ values in Tables 3 and 4). The overall size of the model shrinks as 'concept frequency' declines below the mandatory 'Beaver presence' concept, listed in column two of Tables 5 and 6. Rising connection thresholds tends to disconnect rarer concepts which are consequently orphaned from the aggregated model, as can be seen with

Table 2
Illustration of predictive and observed follow up states for aggregated model version 10, shown in Fig. 2. Transmitters are excluded; Spearman's $\rho = 0.85$, $P = 0.001$.

Concept	Baseline Dynamic Analysis (non-normalised)	Mean observed follow-up state (Scale −3 to +3)
Beaver presence - (mandatory concept)	1.52	1.9
Riparian woodland/vegetation health	0.52	1.1
Nature tourism	0.99	1.7
Beaver acceptance - farmer, landowner, forestry	−0.33	0.5
Beaver acceptance - general public	2.14	1.9
Biodiversity - general	1.42	1.6
Beaver damming and impoundment of water	0.71	0.9
Government policy/finance for nature	0.86	0.8
Wetland - ecosystem health & services	0.64	1.5

Table 3
Correlations between predictive dynamic analysis (DA) and observed states for five versions of the 'permissive' aggregated model. The permissive model is defined by inclusion of the upper two quartiles of concepts by frequency shared between individual mental models, and varied according to additional thresholds for connections based on their frequency (≥2 to ≥6) across contributing mental models.

Aggregated model version	Aggregation: inclusion threshold for number of connections from contributing mental models	Aggregation: n of concepts retained in upper two quartiles	Connections: n	Prediction: DA & observed change; spearman's ρ	P	ρ^2
1	≥2	25	136	0.61	0.013	38 %
2	≥3	23	91	0.61	0.002	38 %
3	≥4	21	60	0.66	0.001	43 %
4	≥5	15	39	0.50	0.058	25 %
5	≥6	12	27	0.64	0.025	41 %

Table 4
Correlations between predictive dynamic analysis (DA) and observed states for five versions of the 'tight' aggregated model. The tight model is defined by inclusion of the upper quartile only of concepts by frequency shared between individual mental models, and varied according to additional thresholds for connections based on their frequency (≥2 to ≥ 6) across contributing mental models.

Aggregated model version	Aggregation: inclusion threshold for number of connections from contributing mental models	Aggregation: n of concepts retained in upper quartile only (*Transmitters)	Connections: n	Prediction: DA & observed change; Spearman's ρ	P	ρ^2
6	≥2	12	56	0.67	0.018	44 %
7	≥3	12	48	0.65	0.023	42 %
8	≥4	12	33	0.81	0.001	65 %
9	≥5	12 (of which *2)	28	0.84	0.001	70 %
10	≥6	12 (of which *3)	21	0.85	0.001	72 %

Table 5
'Permissive' aggregated model, (upper two quartiles of concepts by frequency): concepts are shown with (i) concept frequency across the contributing mental models, (ii) dynamic analysis (DA) predicted scores decided by minimum frequency of connections (≥2 to ≥6) as an inclusion threshold for aggregation, (iii) corresponding mean observed values at three years. Predictive correlations for columns of DA scores and mean observation scores are shown in Table 3.

Concept category (n = 25)	Concept frequency	Predictive value: DA ≥ 2 connections	Predictive value: DA ≥ 3 connections	Predictive value: DA ≥ 4 connections	Predictive value: DA ≥ 5 connections	Predictive value: DA ≥ 6 connections	Mean ObservedValue; 3 years
Beaver presence	31	4.09	3.25	2.82	2.13	1.54	1.89
Biodiversity - general	26	6.26	5.52	3.61	2.96	2.28	1.63
Beaver acceptance - General public	22	6.26	5.06	3.64	2.95	2.54	1.91
Beaver acceptance - Farmer, landowner, forestry	20	0.93	0.19	0.05	0.03	−0.33	0.50
Government policy/ finance for nature	18	3.31	2.73	1.49	0.94	0.87	0.85
Riparian woodland/ vegetation health	16	1.25	1.24	0.86	0.54	0.52	1.09
Wetland - ecosystem health & services	16	3.43	2.73	2.22	1.60	0.64	1.50
Nature tourism	15	3.44	2.22	1.60	1.05	1.02	1.73
Beaver damming and impoundment of water	14	0.77	0.76	0.75	0.74	0.71	0.93
Science, education & knowledge	14	2.71	1.63	1.19	0.69		1.50
Wildlife NGO - effective leadership	14	2.96	1.77	0.62			0.64
Beaver flooding/impact on productive land	13	0.68	0.82	0.54			0.25
Sense of place/ specialness	12	2.25	1.10	1.09	0.53	0.51	2.00
Making space for wilder nature	12	1.78	1.75	1.21	0.60	0.57	1.00
Water retention - upper catchment	12	1.13	1.12	0.40	0.40		1.10
Conflict - public and private property/ amenity	12	0.10	0.35	0.34			−0.30
Beaver acceptance - anglers	11	2.27	1.83	1.21	0.06	−0.40	−0.29
Business generation	11	2.61	1.21	1.18	1.14	0.00	0.70
Holistic enrichment through valuing nature	8	1.67	0.39				1.25
Natural hydrology - wilder river	8	0.87	0.87	0.46			1.33
Flow rate/Problem flooding in lower reaches	8	−0.50	−0.50	−0.50			−1.00
Conflict & distress - natural resource stakeholder	8	−0.09	0.59	0.59			0.57
Water quality	8	0.67					1.57
Fish population health	8	3.13					1.00
Beaver persecution	7	−1.90	−0.10				0.33

increasing frequency towards the foot of Tables 5 and 6.

The overall effect of raising thresholds for both concept inclusion and then connection inclusion is to reduce the number of concepts with their respective activation scores from 25 to 12, noting that in models 9 and 10 shown in Table 4, concepts were excluded from the correlation analysis because they are driver transmitters with fixed activation values.

4. Discussion

4.1. Findings - summary

The findings in the present study provide partial support for H1 and strong support for H2, concerning the question of predictive validity and hence accuracy of same-concept follow-up observer assessments at three years. In addition, this study shows that in aggregated models, tighter thresholds for concept and connection inclusion based on increasingly restrictive criteria of sharedness further improves predictive validity. To our knowledge these contributions are new to the conservation literature. Specifically;

H1. Analysis of predictive and observed FCM activation values for non-aggregated data shows a weak predictive relationship at the level of individual concepts, and a highly variable, inconsistent relationship at the level of individual mental models.

H2. Analysis of predictive and observed FCM activation values for the multi-stakeholder aggregated model shows a strong predictive relationship, statistically significant for nine of ten aggregation inclusion thresholds.

RQ. Investigation of aggregated model concept and connection inclusion thresholds specifying 'sharedness' between constituent mental models shows trade-offs between predictive validity of the aggregated model and model information-richness with respect to numbers of retained concepts and connections.

Table 6
'Tight' aggregated model, (upper quartile of concepts by frequency only): concepts are shown with (i) (concept) frequency across the contributing mental models, (ii) dynamic analysis (DA) predicted scores decided by minimum frequency of connections (≥2 to ≥6) as an inclusion threshold for aggregation, (iii) corresponding mean observed values at three years. Predictive correlations for columns of DA scores and mean observation scores are shown in Table 4.

Concept category (n = 12)	Concept frequency	Predictive value: DA ≥ 2 connections	Predictive value: DA ≥ 3 connections	Predictive value: DA ≥ 4 connections	Predictive value: DA ≥ 5 connections	Predictive value: DA ≥ 6 connections	Mean ObservedValue; 3 years
Beaver presence	31	3.19	3.08	2.28	2.12	1.52	1.89
Biodiversity - general	26	3.29	3.25	2.03	2.02	1.42	1.63
Beaver acceptance - General public	22	4.23	3.53	3.01	2.42	2.14	1.91
Beaver acceptance - Farmer, landowner, forestry	20	0.72	0.15	0.06	0.03	−0.33	0.50
Government policy/ finance for nature	18	3.22	2.67	1.47	0.93	0.86	0.85
Riparian woodland/ vegetation health	16	0.89	0.89	0.55	0.54	0.52	1.09
Wetland - ecosystem health & services	16	2.41	1.72	1.24	1.24	0.64	1.50
Nature tourism	15	1.57	1.07	1.04	1.03	0.99	1.73
Beaver damming and impoundment of water	14	0.76	0.76	0.74	0.74	0.71	0.93
Science, education & knowledge	14	1.88	1.27	1.18	0.69		1.50
Wildlife NGO - effective leadership	14	2.27	1.75	1.18	0.69		0.64
Beaver flooding/impact on productive land	13	0.43	0.82	0.53			0.25

4.2. Implications of the findings

While accurate prediction of expected conservation outcomes is clearly important, mental models have often been viewed as sitting uneasily in this context because they are essentially subjective in nature (Özesmi & Özesmi, 2004). From this perspective, predictive validation of mental models appears questionable. Introducing wisdom of the crowd statistical treatment adds a different and more objective dimension. In this case, the aim is not to show and compare unique individual or sectoral perspectives, but to use combined knowledge to create an objectively validate-able version of the 'real world', including confidence in a 'real' prospective time dimension.

Capacity to make predictions following conservation actions contributes to their credibility and accountability, especially where resources are committed (Game et al., 2018), and outcomes are otherwise contested. Thus, assurance that what is believed will happen is later observed to happen in practice, is a valuable asset. As this and recent studies have shown, aggregation of data consistent with the WOC phenomenon can contribute a new layer of understanding and assurance to conservation planning and interventions. In addition, stakeholder engagement through mental model aggregation gives representative weight to findings which may also increase a sense of democratic legitimacy for policy and decision-makers.

Better model predictions based on aggregation suggest confidence that combined models can be considered both consensual and stable. Additionally, our findings show that more frequent, more highly connected, and presumably important mental model components contribute more strongly to predictive validity at three years. Our most strongly predictive aggregated model (see Tables 2 and 4, model version 10) has twelve interacting concepts showing strong face validity with respect to the basic parameters of beaver reintroduction; items on hydrology and biodiversity, economic inputs and outputs, levels of beaver acceptance amongst various stakeholders, and a key role for conservation NGO leadership.

As increased predictive confidence attaches to connections which survive into high threshold versions of the aggregated model, scholars need to consider whether they wish to focus on a narrower field of predictive accuracy, or retain a softer field of information touching on a wider range of socio-ecological interactions, including concepts and connections which still have the potential to set up scientific hypotheses or management investigations. We suggest that modelling a spectrum of thresholds of decreasing stringency might be valuable where the objective is to broaden a more varied discussion including scenarios featuring less well-validated or foreseeable outcomes.

4.3. The meaning of prediction

There are important limitations to the notion of DA interpreted as a forecast in time. While the iterative calculus used in DA appears at first sight analogous to measured time, what this actually represents is not so clear cut. Stepped models designed by experts to separate out short, medium and longer term futures offer a possible solution (Giordano et al., 2020; Gómez Martín et al., 2020) although these approaches are qualitatively different from perspectives we observe with many non-experts relying on relatively unstructured informal knowledge. Attempting to constrain models by specifying time parameters may increase elicitation stress for some participants, and specifying more 'rules' during elicitation risks distorting perspectives away from the raw participant view towards the pre-conceptions of the researcher, potentially biasing a WOC effect.

While three-year follow-up is modest, it is striking that predictive performance in this case survived Covid restrictions and the legal moment of Brexit, suggesting model robustness. By their nature, both events might be viewed as social rather than ecological drivers. It is possible that a more obviously 'ecological' disaster such as the catastrophic flooding recently experienced in an adjoining county (2014), might have had a more marked impact on predictive validity. Nevertheless, we anticipate that model dynamic analysis will show declining predictive accuracy in the absence of an up-date process. We also anticipate slower decay in model predictive validity in the presence of higher measures of system resilience, with less vulnerability to outcome deviation.

While these questions are unresolved, aggregated mental model dynamic analysis might be thought of as an on-going emergent present, sometimes termed the "long now", (Carpenter, 2002) which can be framed as a span such as 'this year', or 'this decade'. Thinking about model prediction as a description of an extended now is consistent with the instruction to participants in the present study that they consider a five-year span, inferring more nuanced conclusions such as "flow modulation through the watershed is on an increasing trajectory", rather than hazarding more time-bound statements about unfolding but inherently unstable trends and events.

4.4. Some precautions in interpreting the results

The study design invited participants to make judgements about states of the system which they had previously modelled, raising concern about possible confirmation bias. However, dynamic analysis from three years earlier was not available to participants when the models were elicited or subsequently, effectively single blinding the later evaluations. Awareness of system behaviour would have been based on updated participant perceptions rather than the mathematically inferred dynamic analysis.

Purposive recruitment raises the question of possible confounding of predictive accuracy by participant expertise. The 'crowd' recruited into the study and hence into the WOC estimation of system behaviour over time is unlikely to be representative of the wider population. In mental modelling, this is justified as the focus is on accessing knowledge about the social-ecological system, not a survey of public views within which many people may show relatively low levels of interest. The concept of a 'stakeholder' deliberately skews the sample in the present study towards a community of people who have a greater than usual level of engagement with the beaver project and hence beaver interactions with the human and natural environment. Individual and sectoral mental modelling reveals more about beliefs people hold than an assessment of the system as might be measured by external criteria. Mental models are therefore pointers to the kinds of knowledge held by people in a given community, but cannot be generalised further in the way that survey data might be.

In their study, Aminpour et al. (2020) found a paradoxical aggregated model accuracy fall-off when the crowd grew beyond an optimal number, explained by increasingly amplified bias from like-minded subgroups. This finding makes clear that the problem arose less from the size of the crowd than its composition. It is not possible to infer whether the sample size in the present study is too small or too large. Aminpour et al. (2020) addressed the problem by imposing a two-stage analysis, first aggregating stakeholder groups likely to show within-group dependence, and then super-aggregating the resulting sectoral aggregations. In our study, we did not observe obviously co-dependent participants; although possible exceptions included three individuals working for the principal wildlife NGO managing the project. It was notable that these individuals, likely to be viewed as highly expert, chose to highlight dissimilar mental model content reflecting their respective professional focus.

We did not set out to purposively sample for gender, age, or any other social category, focussing instead on affiliation to a relevant range of stakeholder groups. Consequently, it is possible that the recruitment process failed to find important sources of knowledge concealed by inequalities of social profile. We do not know whether more socially inclusive and representative sampling might have further improved predictive validity.

To address some of these potential pitfalls, we suggest that studies combining mental model and wisdom of crowd approaches should consider basic 'rules' of wisdom of the crowd (WOC) theory (Surowiecki, 2005): crowd parameters should (i) reflect the diversity of stakeholders, and (ii) take care to protect independence in modelling *e.g.*, support confidentiality and consider pre-aggregation of larger subgroups at risk of within-group dependence, and (iii) ensure that stakeholders are able to make judgements free from centralised control.

5. Conclusions

The key finding of this paper is that dynamic analysis of broadly-based aggregated mental models showed strong predictive validity for a complex wildlife reintroduction project with important social dimensions. We are not aware that this issue has previously been examined in informal conservation-related stakeholder mental modelling research. We suggest that stronger predictive validity for aggregated mental models can be explained by a 'wisdom of the crowd' (WOC) 'averaging' effect from a diverse group of stakeholders. The study period included major unanticipated intervening societal events, the influence of which cannot be determined from the study itself, however, the strength of the predictive effect in the face of these events is striking. As predictive validity is much weaker for single concepts and extremely inconsistent for individual models, interpretation of dynamic analysis in such cases should be done with greater caution. In summary, it appears that aggregated mental modelling generates stable knowledge from which more confident predictive inferences may be made, at least over short time frames. This observation is offered as an addition to the toolkit of predictive techniques available to conservation policy makers and managers, in the absence of or alongside conventional ecosystem assessments.

Ethical approval

Certified in compliance with the Netherlands Code of Conduct for Research Integrity by Professor Dr. Marcel Verweij, Chair of the Social Sciences Ethics Committee, Wageningen University & Research, 5-12-22.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors wish to thank all our participants, and especially individuals working with Devon Wildlife Trust. This project received no grants or external funding of any kind.

References

Abel, N., Ross, H., & Walker, P. (1998). Mental models in rangeland research, communication and management. *Rangeland Journal, 20*(1), 77–91.

Aminpour, P., Gray, S. A., Jetter, A. A., Introne, J. E., Singer, A., & Arlinghaus, R. (2020). Wisdom of stakeholder crowds in complex social–ecological systems. *Nature Sustainability, 3*(3), 191–199. https://doi.org/10.1038/s41893-019-0467-z

Arlinghaus, R., & Krause, J. (2013). Wisdom of the crowd and natural resource management. *Trends in Ecology and Evolution, 28*(1), 8–11. https://doi.org/10.1016/j.tree.2012.10.009

Auster, R. E., Barr, S. W., & Brazier, R. E. (2021). Improving engagement in managing reintroduction conflicts: Learning from beaver reintroduction. *Journal of Environmental Planning and Management, 64*(10), 1713–1734. https://doi.org/10.1080/09640568.2020.1837089

Auster, R. E., Puttock, A., & Brazier, R. (2020). Unravelling perceptions of Eurasian beaver reintroduction in Great Britain. *Area, 52*(2), 364–375. https://doi.org/10.1111/area.12576

BBC (2014). *Somerset floods crisis: How the story unfolded*. Available at: https://www.bbc.co.uk/news/uk-england-somerset-26157538 (Accessed: 29 October 2022).

Biggs, H., Du Toit, D., Etienne, M., Jones, N., Leitch, A., Lynam, T., … Stone-Jovicich, S. (2008). *A preliminary exploration of two approaches for documenting 'mental models' held by stakeholders in the Crocodile Catchment, South Africa., Water Research Commission*. Available at: https://prodinra.inra.fr/record/30463.

Blewett, A., Jacobs, M., Kok, K., Jones, N., Ogle, A., & Huijbens, E. (2022). Emotionally augmented mental models, connectivity and beaver reintroduction in Southwest

England. *Ecology and Society, 27*(1), Article 33. https://doi.org/10.5751/ES-12823-270133
Blewett, A., Jacobs, M., Kok, K., Jones, N., & Ogle, S. (2021). Stakeholder mental model analysis supports focused conservation policy and actions for Eurasian beaver (*Castor fiber*) reintroduction. *Journal for Nature Conservation, 64*(March), Article 126064. https://doi.org/10.1016/j.jnc.2021.126064
Brazier, R. E., Elliott, M., Andison, E., Auster, R. E., Bridgewater, S., Burgess, P., … Vowles, A. (2020) *River Otter Beaver Trial science and evidence report*. Available at: http://www.exeter.ac.uk/media/universityofexeter/research/microsites/creww/riverottertrial/ROBT_Science_and_Evidence_Report_2020_(ALL).pdf.
Brazier, R. E., Puttock, A., Graham, H., Auster, R. E., Davies, K. H., & Brown, C. M. L. (2021). Beaver: Nature's ecosystem engineers. *Wiley Interdisciplinary Reviews: Water, 8*(1), 1–29. https://doi.org/10.1002/wat2.1494
Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior, 22*(2), 195–202. https://doi.org/10.1002/job.82
Carpenter, S. R. (2002). Ecological futures: Building an ecology of the long now. *Ecology, 83*(8), 2069–2083.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155
Conservation Evidence Base, Department of Zoology University of Cambridge, U.K. (no date). Available at: https://www.conservationevidence.com (Accessed: 2 March 2021).
Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.
Estupiñán Ricardo, Coka Flores, Erás Diaz, & Pérez Teruel, K. (2020). An exploration of wisdom of crowds using neutrosophic cognitive maps. Neutrosophic Sets and Systems, *37*(Special Issue: Impact of neutrosophy in solving the Latin American's social problems), 8–15.
Galton, F. (1907). *Vox populi. Nature, 75*, 450–451.
Game, E. T., Bremer, E. L., Calvache, L., Moreno, P. H., Vargas, A., Rivera, B., & Rodriquez, L. M. (2018). Fuzzy models to inform social and environmental indicator selection for conservation impact monitoring. *Conservation Letters*. https://doi.org/10.1111/conl.12338
Giordano, R., Pluchinota, I., Pagano, A., Scrieciu, A., & Nanu, F. (2020). Enhancing nature-based solutions acceptance through stakeholders' engagement in co-benefits identification and trade-offs analysis. *Science of the Total Environment, 713*. https://doi.org/10.1016/j.scitotenv.2020.136552
Gómez Martín, E., Giordano, R., Pagano, A., van der Kaur, P., & Máñez Costa, M. (2020). Using a system thinking approach to assess the contribution of nature based solutions to sustainable development goals. *Science of the Total Environment, 738*. https://doi.org/10.1016/j.scitotenv.2020.139693
Gray, S., Chan, A., Clark, D., & Jordan, R. (2012). Modeling the integration of stakeholder knowledge in social-ecological decision-making: Benefits and limitations to knowledge diversity. *Ecological Modelling, 229*, 88–96. https://doi.org/10.1016/j.ecolmodel.2011.09.011
Gray, S. A., Zanre, E., & Gray, S. R. J. (2014). Fuzzy cognitive maps as representations of mental models and group beliefs. *Intelligent Systems Reference Library*, 29–48. https://doi.org/10.1007/978-3-642-39739-4_2
Jetter, A. J., & Kok, K. (2014). Fuzzy Cognitive Maps for futures studies—A methodological assessment of concepts and methods. *Futures, 61*, 45–57. https://doi.org/10.1016/j.futures.2014.05.002
Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences, 107*(43), 18243–18250. https://doi.org/10.1073/pnas.1012933107
Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental models: An interdisciplinary synthesis of theory and methods. *Ecology and Society, 16*(1), Article 46. https://doi.org/10.5751/ES-03802-160146
Kok, K. (2009). The potential of Fuzzy Cognitive Maps for semi-quantitative scenario development, with an example from Brazil. *Global Environmental Change, 19*(1), 122–133. https://doi.org/10.1016/j.gloenvcha.2008.08.003
Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies*. https://doi.org/10.1016/S0020-7373(86)80040-2
Liao, Y., Hsieh, H.-L., Xu, S., Zhong, Q., Lei, J., Liang, M., & Kwan, B. K. Y. (2019). Wisdom of Crowds reveals decline of Asian horseshoe crabs in Beibu Gulf, China. *Oryx, 53*(2), 222–229. https://doi.org/10.1017/S003060531700117X
Moon, K., Guerrero, A. M., Adams, V. M., Biggs, D., Blackman, D. A., Craven, L., & Ross, H. (2019). Mental models for conservation research and practice. *Conservation Letters, 12*(3), 1–11. https://doi.org/10.1111/conl.12642
Nyhus, P. J. (2016). Human-wildlife conflict and coexistence. *Annual Review of Environment and Resources*. https://doi.org/10.1146/annurev-environ-110615-085634
Obiedat, M., & Samarasinghe, S. (2013). Fuzzy representation and aggregation of fuzzy cognitive maps. In *Proceedings - 20th international congress on modelling and simulation, MODSIM 2013*, (December), 684–690. doi: 10.36334/modsim.2013.c2.obiedat.
Obiedat, M., & Samarasinghe, S. (2016). A novel semi-quantitative Fuzzy Cognitive Map model for complex systems for addressing challenging participatory real life problems. *Applied Soft Computing, 48*, 91–110. https://doi.org/10.1016/j.asoc.2016.06.001
Olazabal, M., Neumann, M. B., Foudi, S., & Chiabai, A. (2018). Transparency and reproducibility in participatory systems modelling: The case of fuzzy cognitive mapping. *Systems Research and Behavioral Science, 35*(6), 791–810. https://doi.org/10.1002/sres.2519
Özesmi, U., & Özesmi, S. L. (2004). Ecological models based on people's knowledge: A multi-step fuzzy cognitive mapping approach. *Ecological Modelling, 176*(1–2), 43–64. https://doi.org/10.1016/j.ecolmodel.2003.10.027
Rouse, W. B., & Morris, N. M. (1986). On looking into the black box. Prospects and limits in the search for mental models. *Psychological Bulletin, 100*(3), 349–363. https://doi.org/10.1037/0033-2909.100.3.349
Surowiecki, J. (2005). *The Wisdom of Crowds: Why the Many are Smarter than the Few*. New York, NY: Doubleday.
Verkerk, P. J., Sánchez, A., Libbrecht, S., Broekman, A., Bruggeman, A., Daly-Hassen, H., … Zoumides, C. (2017). A participatory approach for adapting river basins to climate change. *Water (Switzerland), 8*(12), 1–16. https://doi.org/10.3390/w9120958
Whitmore, N. (2016). Harnessing local ecological knowledge for conservation decision making via Wisdom of Crowds: The case of the Manus green tree snail *Papustyla pulcherrima*. *Oryx, 50*(4), 684–692. https://doi.org/10.1017/S0030605315000526
Wildenberg, M., Bachhofer, M., Adamescu, M., de Blust, G., Diaz-Delgadod, R., Isak, K. G. Q., … Riku, V. (2010). Linking thoughts to flows – Fuzzy cognitive mapping as tool for integrated landscape modelling. In *Proceedings of the 2010 International Conference on Integrative Landscape Modelling - Linking Environmental, Social and Computer Sciences*, February 3-5. Montpellier, France.