Bayesian Information Fusion for Consumer Understanding and Product Innovation

data&modelingsciences

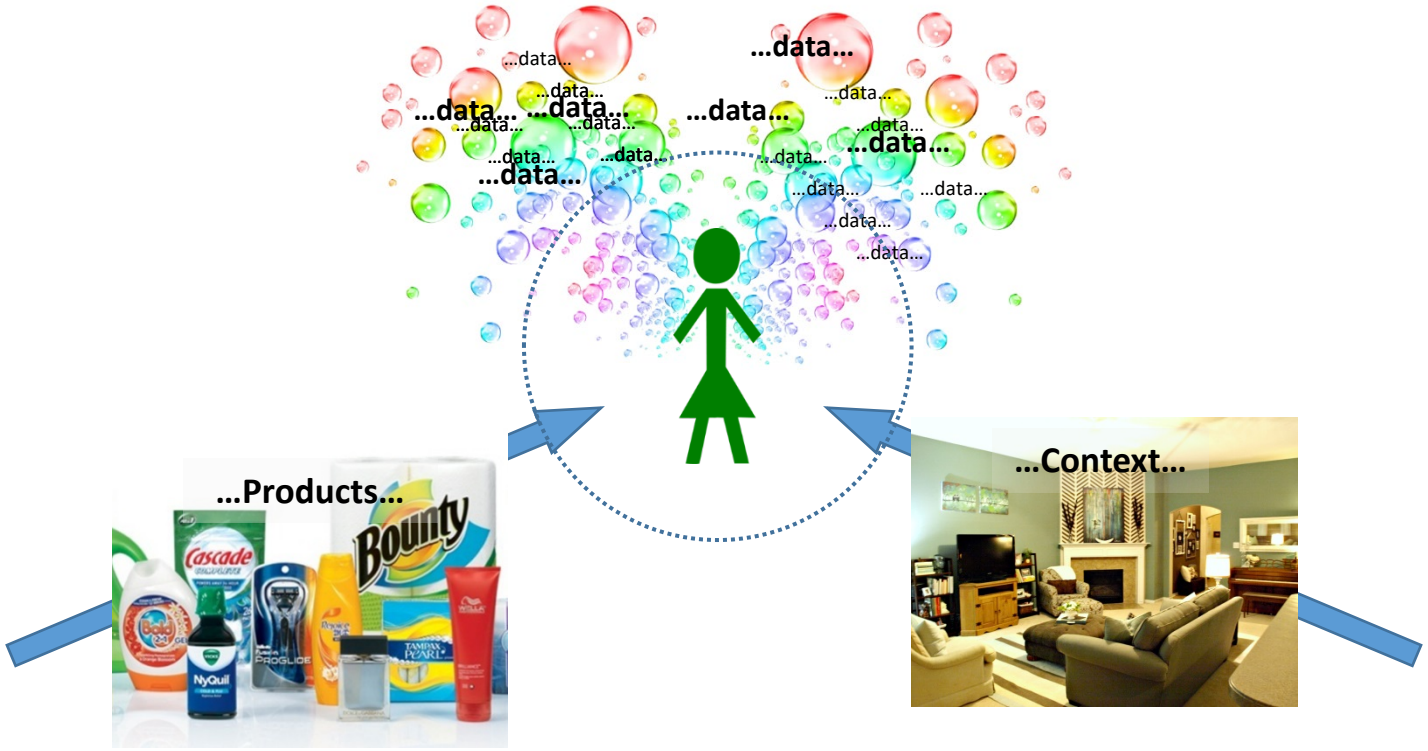
Yong Zhang, Ph.D. Michael L. Thompson, Ph.D.
Sabine Metzger-Groom, Ph.D. Amy Phillips

Procter & Gamble, Corporate Function R&D, DMS & FHC
(BayesiaLab Conference, 10/10/2019)

Overview

- Bayesian Information Fusion
- Assumptions and Methods
- Validity Check Using Simulated Data
- P&G Case Study: Link Overall Rating (OAR) of study A to OAR of Study B
- Summary

Consumer's Usage Experience



Bayesian Information Fusion Model

Quantitatively Extracts & Summarizes
Information from Different Types of Studies

Consistence within & across Studies / Data

Qualitative Summary
Conclusion
A, B &C

Cross Inference among Studies / Data

Study A
Findings a1, a2

Study C
Findings c1, c2

Study B
Findings b1, b2

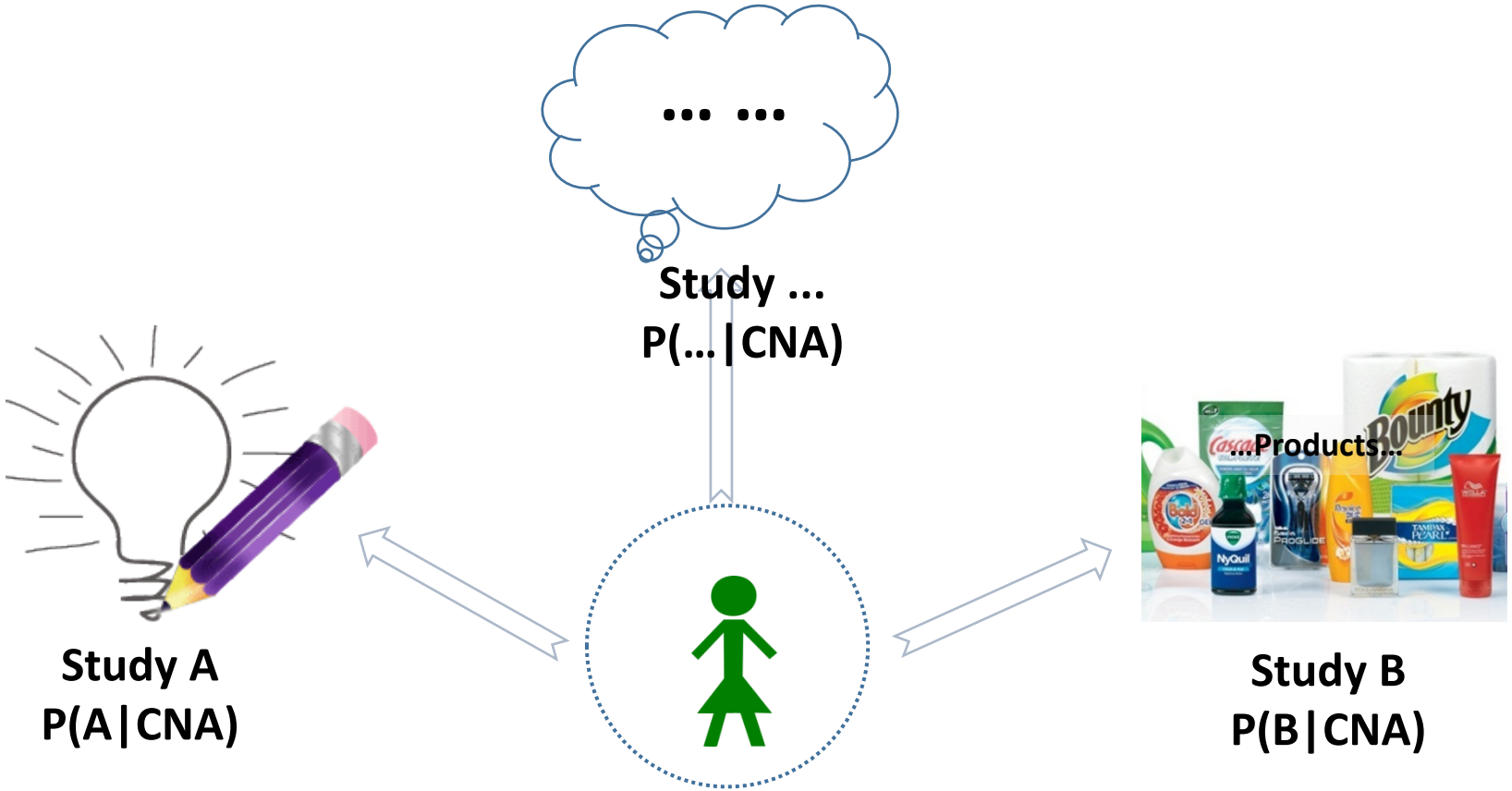
Assumptions

- Studies in different contexts share a set of “common variables” such as consumer needs and attitudes

- This set of “common variables” can be consistently measured across contexts; i.e., they are independent of contexts

Allenby G. et al. (2014), Rossi P. et al (2005)

Methods Intuition



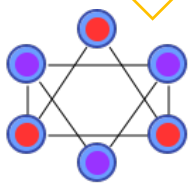
$$P(CNA, A, B) = P(CNA) \times P(A|CNA) \times P(B|CNA) \times P(\ldots|CNA)$$

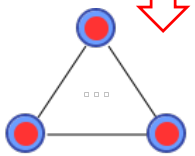
(Chain Rule of Bayesian Belief Network (BBN))

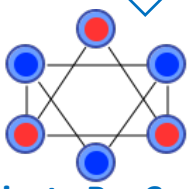
Methods Implementation: BBN

Study A **Consumer** **Study B**

	aOAR	aVar	...	Need	Attitude	...	bOAR	bVar	...
Consumer 1	✓	✓	✓	✓	✓	✓	?	?	?
Consumer 2	✓	✓	✓	✓	✓	✓	?	?	?
...	✓	✓	✓	✓	✓	✓	?	?	?
Consumer N	✓	✓	✓	✓	✓	✓	?	?	?
Consumer 1	?	?	?	✓	✓	✓	✓	✓	✓
Consumer 2	?	?	?	✓	✓	✓	✓	✓	✓
...	?	?	?	✓	✓	✓	✓	✓	✓
Consumer M	?	?	?	✓	✓	✓	✓	✓	✓

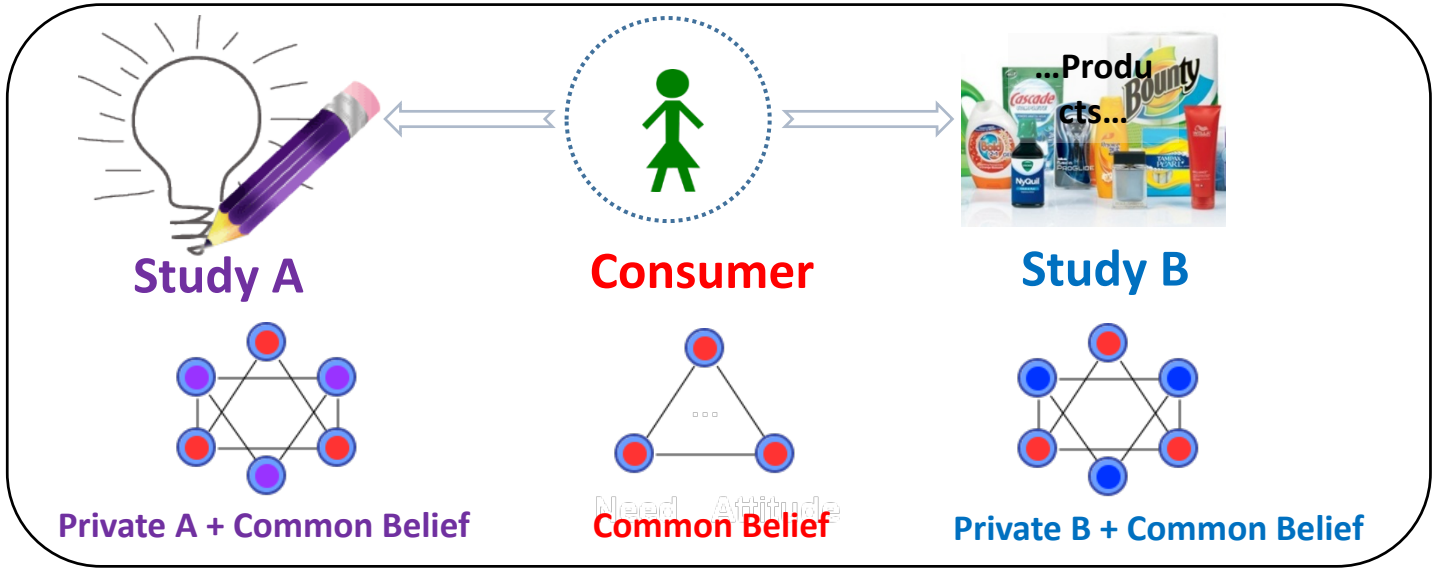


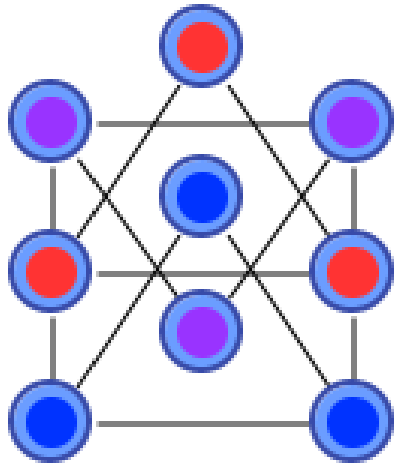




Step2: Private A + Common Belief **Step1: Common Belief** **Step2: Private B + Common Belief**

Methods Implementation: BBN



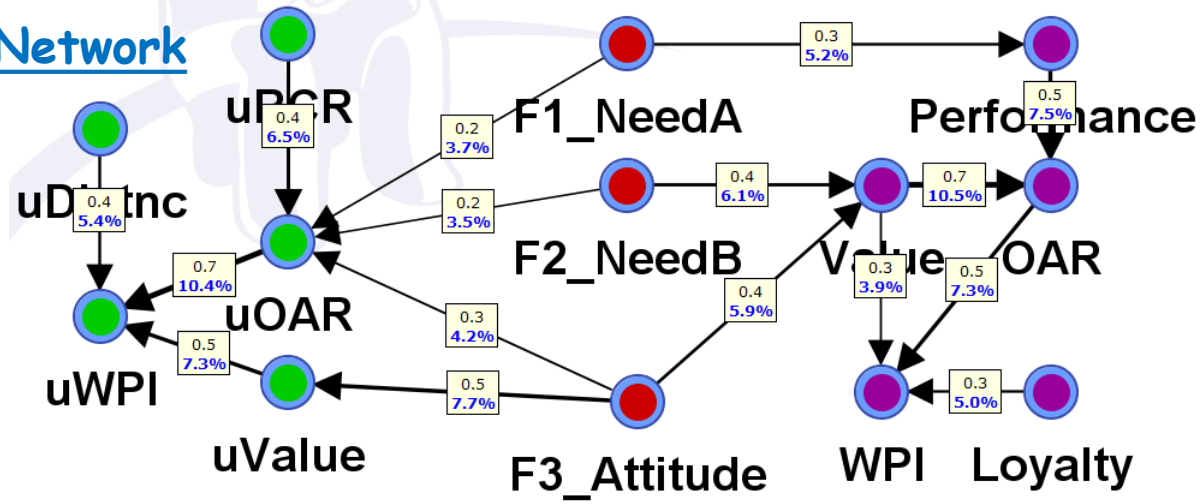


Koller D. et al. (1997),
Xiang Y. et al. (2003),
Weber P. et al. (2006)

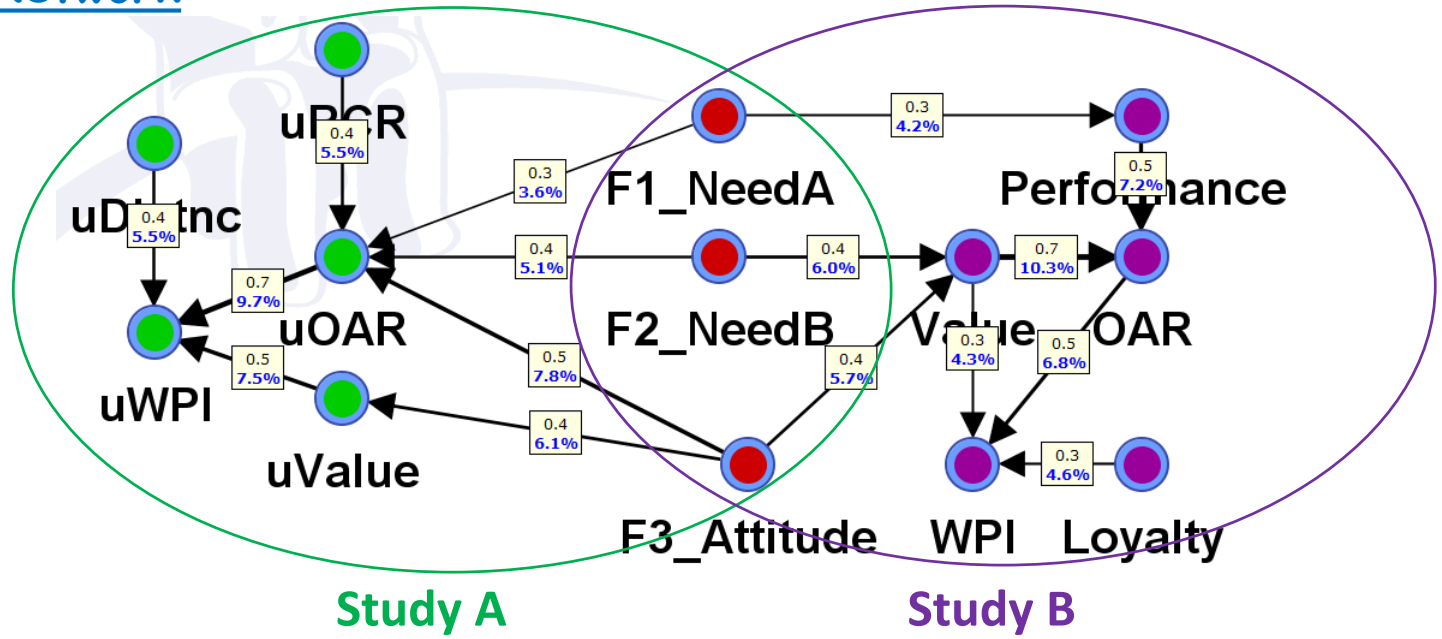
Step3: BBN Model to Link Study A and Study B

Validation on Simulated Data: Structure

Ground Truth Network

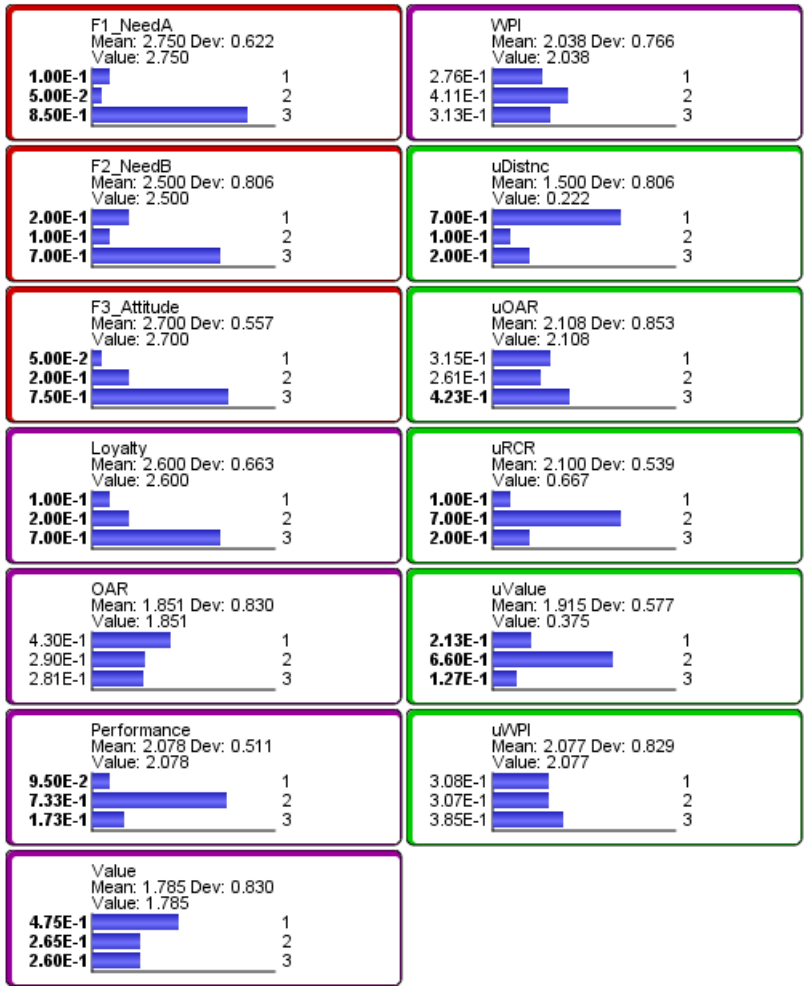


Recovered Network

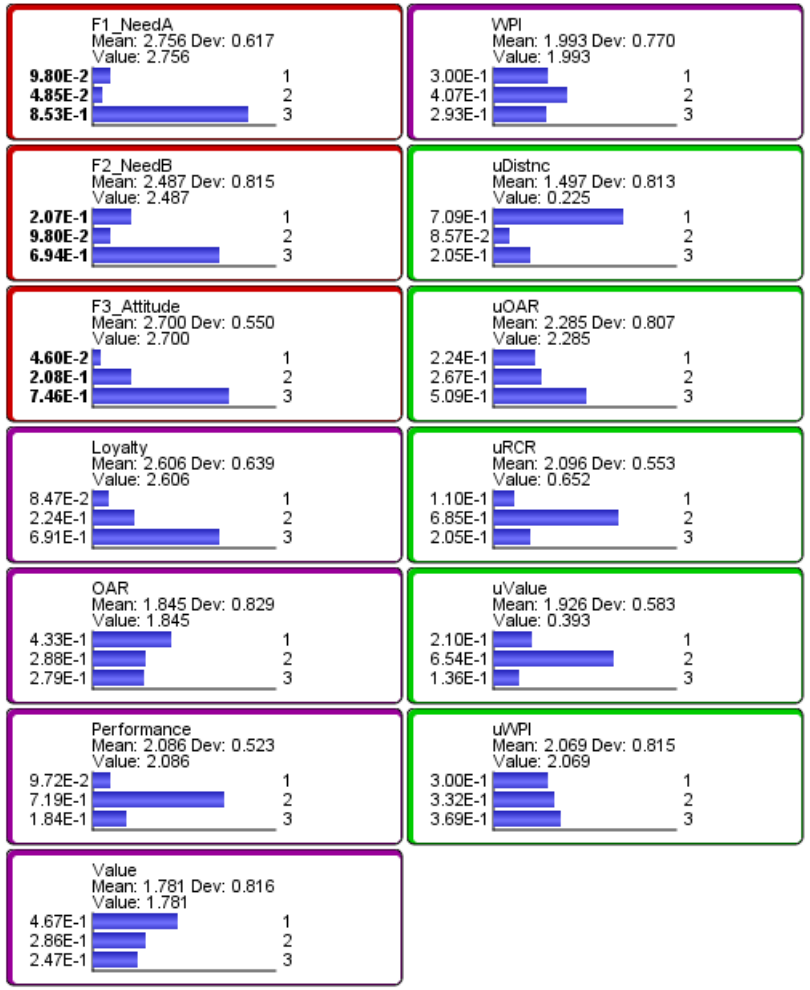


Validation on Simulated Data: Distribution

Ground Truth Network



Recovered Network



Validation on Simulated Data: Prediction

Ground Truth Network

Database: Test Samples

Node	Comment	Overall Precision	Mean Precision	R2	RMSE	NRMSE	Overall LogLoss
F1_NeedA		90.7500%	77.2060%	0.6218	0.3915	19.5727%	0.2283
F2_NeedB		82.0000%	81.3461%	0.4746	0.5918	29.5908%	0.4191
F3_Attitude		93.0000%	87.4180%	0.7280	0.2943	14.7128%	0.2109
Loyalty		77.5000%	56.2728%	0.2230	0.5766	28.8304%	0.5567
OAR		84.2500%	82.7121%	0.6959	0.4593	22.9643%	0.3612
Performance		85.2500%	74.9432%	0.5709	0.3549	17.7465%	0.3150
Value		88.7500%	87.8210%	0.7729	0.3864	19.3200%	0.2829
WPI		78.0000%	77.0025%	0.4404	0.5733	28.6645%	0.6774
uDistnc		79.5000%	52.8701%	0.4602	0.3185	31.8508%	0.5513
uOAR		83.5000%	83.5275%	0.6776	0.4742	23.7116%	0.3618
uRCR		79.2500%	77.2567%	0.4630	0.3434	34.3381%	0.5120
uValue		87.2500%	71.8872%	0.7657	0.2286	22.8639%	0.2877
uWPI		83.7500%	83.4472%	0.6061	0.5120	25.5982%	0.5476

Database: Learning Samples

Node	Comment	Overall Precision	Mean Precision	R2	RMSE	NRMSE	Overall LogLoss
F1_NeedA		92.9375%	80.9838%	0.6174	0.3811	19.0562%	0.1937
F2_NeedB		81.0000%	77.5882%	0.4499	0.6052	30.2603%	0.4363
F3_Attitude		93.5625%	87.0743%	0.7314	0.2831	14.1536%	0.1734
Loyalty		76.0000%	55.1183%	0.2046	0.5617	28.0851%	0.5833
OAR		83.3125%	81.8034%	0.6648	0.4821	24.1035%	0.3948
Performance		87.1875%	74.1724%	0.5335	0.3513	17.5642%	0.3225
Value		87.6875%	86.3236%	0.7329	0.4246	21.2282%	0.3324
WPI		76.1250%	74.6419%	0.3784	0.6053	30.2646%	0.7077
uDistnc		80.0000%	54.7091%	0.4535	0.3082	30.8228%	0.5337
uOAR		82.7500%	83.0522%	0.6676	0.4875	24.3738%	0.3761
uRCR		78.7500%	77.1516%	0.4945	0.3382	33.8223%	0.4982
uValue		84.9375%	68.0340%	0.6975	0.2674	26.7431%	0.3382
uWPI		81.5625%	81.3589%	0.5242	0.5678	28.3896%	0.6000

Recovered Network

Database: Test Samples

Node	Comment	Overall Precision	Mean Precision	R2	RMSE	NRMSE	Overall LogLoss
F1_NeedA		80.0000%	58.1624%	0.1519	0.5875	29.3762%	0.4960
F2_NeedB		62.5000%	53.3784%	0.1064	0.8708	43.5423%	0.9641
F3_Attitude		79.0000%	61.5930%	0.3247	0.4902	24.5120%	0.5583
Loyalty		77.0000%	50.2547%	0.2221	0.5508	27.5391%	0.5761
OAR		83.5000%	81.9935%	0.6803	0.4742	23.7084%	0.3598
Performance		88.5000%	73.8170%	0.4867	0.3359	16.7929%	0.3045
Value		90.5000%	89.4792%	0.7608	0.3802	19.0094%	0.2600
WPI		79.0000%	76.0212%	0.3383	0.5942	29.7080%	0.6730
uDistnc		79.2500%	54.7467%	0.4694	0.3310	33.1042%	0.5800
uOAR		66.5000%	62.0655%	0.3984	0.6967	34.8344%	1.0815
uRCR		68.5000%	43.6039%	0.0096	0.6054	60.5431%	1.3238
uValue		83.5000%	70.4191%	0.6408	0.2974	29.7403%	0.3583
uWPI		81.7500%	81.8528%	0.5533	0.5276	26.3806%	0.6149

Database: Learning Samples

Node	Comment	Overall Precision	Mean Precision	R2	RMSE	NRMSE	Overall LogLoss
F1_NeedA		75.6250%	57.3098%	0.2432	0.6222	31.1106%	0.5931
F2_NeedB		65.5000%	54.7096%	0.1124	0.8331	41.6565%	0.9817
F3_Attitude		80.1250%	64.7297%	0.3679	0.4721	23.6074%	0.6072
Loyalty		76.9375%	51.3925%	0.2022	0.5679	28.3950%	0.5872
OAR		83.1875%	81.7405%	0.6617	0.4833	24.1635%	0.4035
Performance		87.0000%	75.3483%	0.5737	0.3475	17.3752%	0.3226
Value		87.8125%	86.3180%	0.7364	0.4251	21.2528%	0.3454
WPI		75.6875%	74.6288%	0.3997	0.6026	30.1299%	0.7222
uDistnc		79.7500%	54.3912%	0.4364	0.3082	30.8203%	0.5229
uOAR		66.6875%	62.8869%	0.4011	0.6816	34.0785%	1.1663
uRCR		66.1875%	42.1197%	0.0227	0.6133	61.3282%	1.3432
uValue		85.7500%	74.6070%	0.7344	0.2475	24.7489%	0.3282
uWPI		82.0625%	81.8274%	0.5375	0.5633	28.1637%	0.5846

Validation on Simulated Data: Inference*

	Ground Truth Network				Recovered Network			
	Precision		Reliability		Precision		Reliability	
	OAR	WPI	OAR	WPI	OAR	WPI	OAR	WPI
Study A Only	60.7%	42.1%	60.0%	39.0%	60.7%	42.1%	60.0%	39.0%
Study B Only	82.6%	76.0%	82.5%	76.2%	83.2%	75.7%	83.4%	75.8%
All Data	82.6%	76.0%	82.5%	76.2%	83.2%	75.7%	83.4%	75.8%
Random #	0.00	1.5×10^{-9}	0.00	2.2×10^{-5}	0.00	1.5×10^{-9}	0.00	2.2×10^{-5}

* Basesize is 1000, independent from 2000 learning and develop data points
# Achieve the same Precision/Reliability as "Study A Only" by randomness

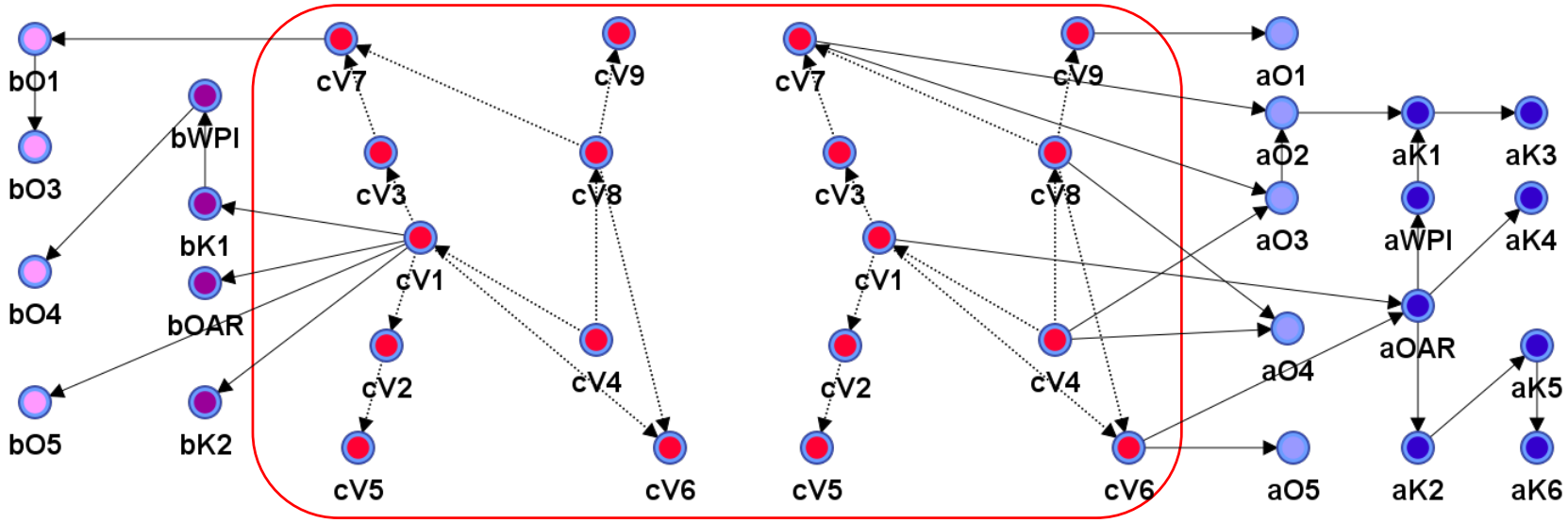
P&G Case Study: Procedure

- **Step 1: Pick “Common Variables” shared across studies**
- **Step 2: Build a BBN subnetwork on “Common Variables”**
- **Step 3: Build BBN subnetworks on data of studies A and B, separately**
- **Step 4: Link BBN subnetworks together**

Represent knowledge using BBN network

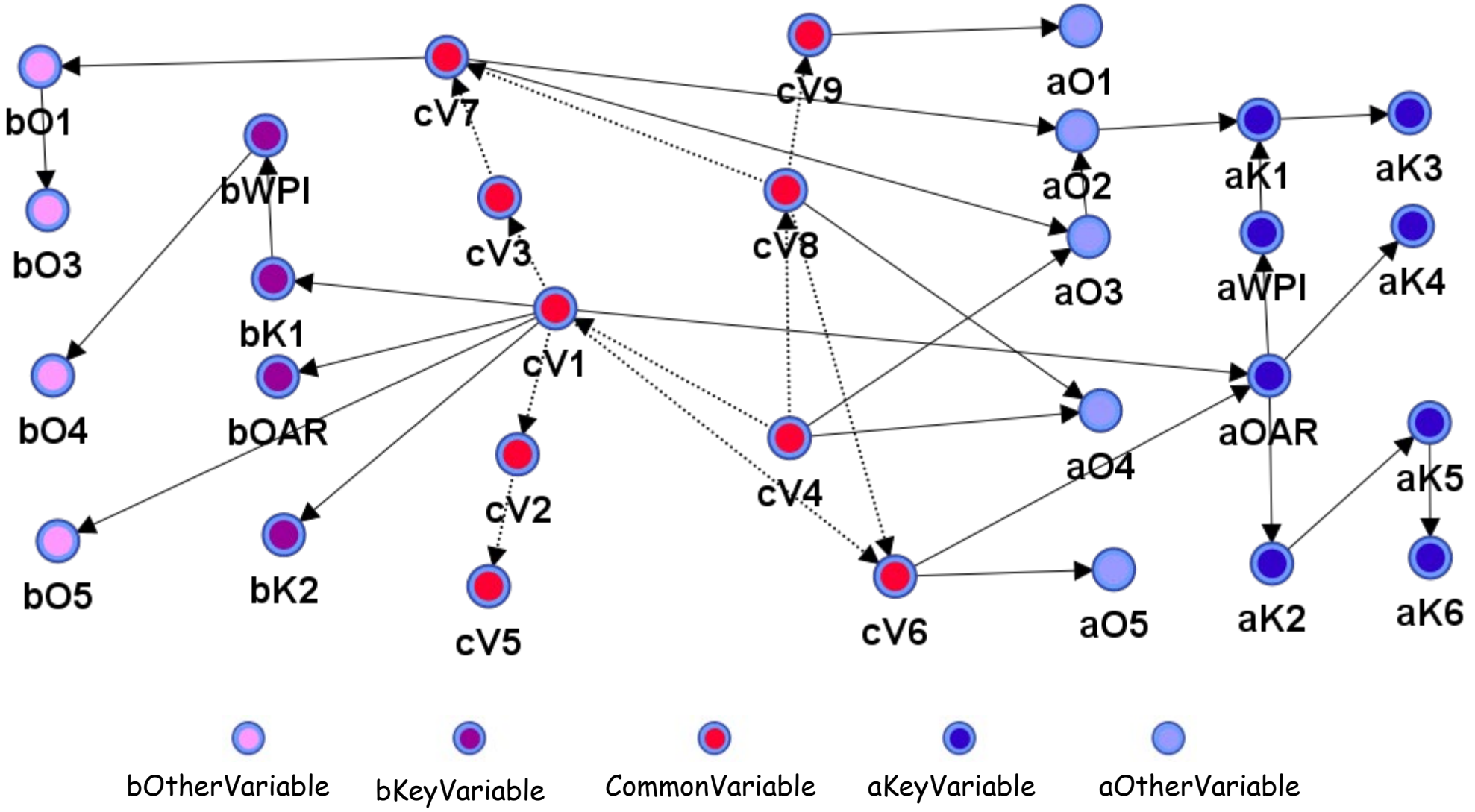
Subnetwork representing knowledge learnt from study B

Subnetwork representing knowledge learnt from study A



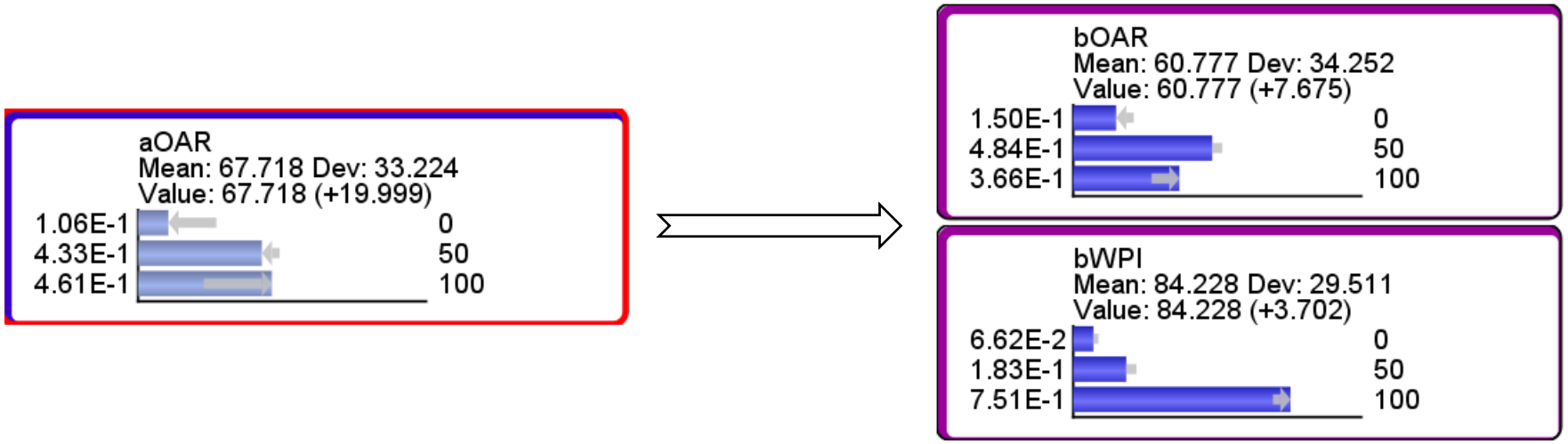
Subnetwork representing common belief learnt from variables shared by Studies A and B

Overall knowledge learnt from studies A and B



Cross Inference: Link OAR of context A to OAR of context B

Q: What is the response of bOAR in context b if we observe a 20% boost of aOAR in context a? Assume this aOAR boost is achieved through improvement in product benefit cV1.



A: bOAR in context b will be boosted by 7.7% on average. This will make more users in context b thinking the corresponding product is excellent (i.e., top box OAR), and increase purchase intent by 3.7% on average.

Consistence across Contexts/Studies

Are total effects on overall rating consistent across studies?

Product Benefits	Study B STE	Study A STE	Rank Product Benefit-Study B	Rank Product Benefit-Study A
cV1	0.63	0.61	1	1
cV2	0.49	0.46	2	3
cV3	0.45	0.41	3	4
cV4	0.44	0.36	4	5
cV5	0.39	0.32	5	6
cV6	0.35	0.52	6	2
cV7	0.32	0.30	7	7
cV8	0.27	0.30	8	8
cV9	0.21	0.18	9	9

Cosine similarity between two ranks: 96.5%.

Summary

- Bayesian Information Fusion Model can be used to integrate different studies into a BBN network. This network represents overall knowledge of complex domain and different contexts;
- The whole modeling process can be implemented using existing software BayesiaLab;
- Omni-direction inference can be conducted within a single study, and across multiple studies;
- Validity was checked on simulated data and a P&G dataset.