Online appendix for:

The necessity of discretion:
A behavioral evaluation of bottom-up implementation theory

Journal of Public Administration Research and Theory

Authors:

First, corresponding author:
Dr. Eva Thomann
University of Exeter
Department of Politics
Amory building, Rennes Drive
EX4 4RJ Exeter
United Kingdom
E-Mail: e.thomann@exeter.ac.uk

Second author:
Nadine van Engen
Erasmus University Rotterdam
Department of Public Administration
Postbus 1738
3000 DR Rotterdam
The Netherlands
vanengen@essb.eur.nl

Third author:
Prof. Dr. Lars Tummers
Utrecht University
Utrecht School of Governance
Bijlhouwerstraat 6
3511 ZC Utrecht
The Netherlands
E-Mail: l.g.tummers@uu.nl

Online Appendix I: Supplementary material

Table B1. Measurement and descriptive statistics of indicator variables, datasets 1 and 2

Set	*Survey question*	*Min*	*Max*	*Mean*	*Standard deviation*	*Skew*	*% missing*	*% cases with set membership > 0.5*
High implementation willingness W	W1: I intend to try to convince employees of the benefits the DRG policy will bring	1	5	1.93	0.86	0.69	19.2	43.3
	W1: I intend to try to convince colleagues of the benefits of government policies	*1*	*5*	*2.79*	*0.97*	*0.01*	*2.0*	*72.4*
	W2: I intend to put effort into achieving the goals of the DRG policy	1	5	2.54	1.05	0.05	18.7	
	W2: I intend to put effort into achieving the goals of government policies	*1*	*5*	*3.35*	*0.95*	*-0.51*	*1.4*	
	W3: I intend to reduce resistance among employees regarding the DRG policy	1	5	2.27	0.93	0.26	19.4	
	W3: I intend to reduce resistance among colleagues regarding government policies	*1*	*5*	*2.93*	*0.99*	*-0.11*	*2.4*	
	W4: I intend to make time to implement the DRG policy	1	5	2.64	1.02	-0.09	20.1	
	W4: I intend to make time to implement government policies	*1*	*5*	*3.33*	*0.98*	*-0.51*	*1.6*	
	W5: I intend to put effort in, in order to implement the DRG policy successfully	1	5	2.98	1.02	-0.51	19.0	
	W5: I intend to put effort in, in order to implement government policies successfully	*1*	*5*	*3.59*	*0.89*	*-0.79*	*1.0*	
High strategic powerfulness SP	SP1: On a national level, professionals could influence how the policy was set up[1]	1	5	3.43	0.99	-0.21	17.8	45.1
	SP1: On a national level, professionals can influence how policies are set up[1]	*1*	*5*	*3.69*	*0.91*	*-0.70*	*0.6*	*56.3*
	SP2: Professionals, through their professional associations, actively helped to think with the design of the policy[1]	1	5	3.12	0.99	0.12	17.6	
	SP2: Professionals, through their professional associations, actively help in drawing up the design of government policies[1]	*1*	*5*	*2.70*	*0.99*	*0.31*	*1.0*	
	SP3: In my opinion, professionals had too little power to influence the policy	1	5	4.00	0.91	-0.98	16.5	
	SP3: In my opinion, professionals have too little power to influence government policies	*1*	*5*	*4.21*	*0.83*	*-1.15*	*0.2*	
	SP4: Politicians did not, during the design of the policy, listen to the professionals at all	1	5	3.63	0.93	-0.23	17.0	
	SP4: Politicians, during the design of policies, do not listen to professionals at	*1*	*5*	*3.98*	*0.93*	*-0.63*	*0.4*	

Set	*Survey question*	*Min*	*Max*	*Mean*	*Standard deviation*	*Skew*	*% missing*	*% cases with set membership > 0.5*
	all							
	SP5: We professionals were completely powerless during the introduction of the policy	1	5	3.42	1.05	-0.11	16.7	
	SP5: We, as professionals, are completely powerless during the introduction of government policies	*1*	*5*	*3.93*	*0.99*	*-0.68*	*0.4*	
	SP6: Professionals could not at all influence the development of the policy at the national level (Minister and Ministry of X, National Government)	1	5	3.50	0.99	-0.19	17.0	
	SP6: Professionals cannot influence the development of policies at the national level (Minister and Ministry of X, National Government)	*1*	*5*	*3.88*	*0.97*	*-0.65*	*0.1*	
High tactical powerfulness TP	TP1: In my organization, especially professionals could decide how the policy was being implemented[1]	1	5	3.78	1.00	-0.82	19.5	44.8
	TP1: In my organization, it is especially professionals who decide how government policies are implemented [1]	*1*	*5*	*3.33*	*1.03*	*-0.37*	*0.2*	*66.3*
	TP2: In my organization, professionals have – by means of working groups or meetings – taken part in decisions on the execution of the policy[1]	1	5	3.58	1.06	-0.42	21.6	
	TP2: In my organization, professionals - through working groups or meetings - take part in decisions on executing government policies[1]	*1*	*5*	*2.85*	*1.04*	*0.31*	*0.1*	
	TP3: The management of my organization should have involved the professionals far more in the execution of the policy	1	5	3.51	0.99	-0.38	23.0	
	TP3: The management of my organization should involve professionals far more in the execution of government policies	*1*	*5*	*3.95*	*0.96*	*-0.78*	*0.5*	
	TP4: Professionals were not listened to over the introduction of the policy in my organization	1	5	3.18	1.05	-0.02	22.6	
	TP4: Professionals are not listened to during the introduction of government policies in my organization	*1*	*5*	*2.91*	*1.12*	*0.35*	*0.1*	
	TP5: In my organization, professionals could take part in conversations regarding the execution of the policy[1]	1	5	3.32	1.06	-0.21	22.4	
	TP5: In my organization, professionals take part in conversations regarding the execution of government policies [1]	*1*	*5*	*2.73*	*1.02*	*0.48*	*0.1*	
	TP6: I and my fellow professionals were completely powerless in the introduction of the policy in my organization	1	5	3.37	1.14	-0.25	20.4	

Set	*Survey question*	*Min*	*Max*	*Mean*	*Standard deviation*	*Skew*	*% missing*	*% cases with set membership > 0.5*
	TP6: I and my fellow professionals are completely powerless during the introduction of government policies in my organization	*1*	*5*	*3.01*	*1.16*	*0.21*	*0.2*	
High operational powerfulness OP	OP1: I have freedom to decide how to use the policy[1]	1	5	3.59	1.10	-0.60	16.1	56.7
	OP1: Generally, I have freedom to decide how to use government policies[1]	*1*	*5*	*3.21*	*0.98*	*-0.18*	*0.5*	*56.3*
	OP2: While working with the policy, I can be in keeping with the client's needs[1]	1	5	3.71	1.05	-0.56	17.0	
	OP2: Generally, when working with government policies, I can be in keeping with clients' needs[1]	*1*	*5*	*3.26*	*0.92*	*-0.18*	*0.6*	
	OP3: Working with the policy feels like a harness in which I cannot easily move	1	5	3.46	1.11	-0.28	16.2	
	OP3: Generally, working with government policies feels like a harness in which I cannot easily move	*1*	*5*	*3.24*	*1.06*	*-0.06*	*0.6*	
	OP4: When I work with the policy, I have to adhere to tight procedures	1	5	3.44	1.04	-0.28	15.9	
	OP4: Generally, when working with government policies, I have to adhere to tight procedures	*1*	*5*	*3.37*	*0.93*	*-0.19*	*1.1*	
	OP5: While working with the policy, I cannot sufficiently tailor it to the needs of my clients	1	5	3.19	1.15	0.03	16.1	
	OP5: Generally, government policies allow me to sufficiently tailor them to the needs of my clients	*1*	*5*	*3.42*	*1.03*	*-0.17*	*0.7*	
	OP6: While working with the policy, I can make my own judgments[1]	1	5	2.97	1.03	0.25	15.8	
	OP6: Generally, government policies allow me to make my own judgments [1]	*1*	*5*	*3.01*	*0.93*	*0.13*	*0.8*	
High societal meaningfulness SM	SM1: I think that the policy, in the long term, will lead to goal 1[1,2]	1	5	3.69	1.04	-0.47	15.8	27.2
	SM1: In general, I think that government policy in the long term will lead to socially relevant goal A [1]	*1*	*5*	*3.33*	*1.04*	*-0.8*	*0.0*	*33.6*
	SM2: I think that the policy, in the short term, will lead to goal 1[1,2]	1	5	3.99	0.86	-0.72	15.9	
	SM2: In general, I think that government policy in the short term will lead to socially relevant goal A [1]	*1*	*5*	*3.6*	*0.92*	*-0.34*	*0.0*	
	SM3: I think that the policy has already led to goal 1[1,2]	1	5	4.02	0.85	-0.72	16.0	
	SM3: In general, I think that government policy has already led to socially relevant goal [1]	*1*	*5*	*3.69*	*0.90*	*-0.47*	*0.1*	
	SM4: Overall, I think that the policy leads to goal 1[1,2]	1	5	3.80	1.02	-0.55	15.9	
	SM4: Overall, I think that government policy leads to socially relevant goal A	*1*	*5*	*3.43*	*1.00*	*-0.19*	*0.0*	

Set	*Survey question*	*Min*	*Max*	*Mean*	*Standard deviation*	*Skew*	*% missing*	*% cases with set membership > 0.5*
	1[1,2]							
	SM5: I think that the policy, in the long term, will lead to goal 2[1,2]	1	5	3.82	1.01	-0.56	16.0	
	SM6: I think that the policy, in the short term, will lead to goal 2[1,2]	1	5	4.00	0.88	-0.74	16.0	
	SM7: I think that the policy has already led to goal 2[1,2]	1	5	4.03	0.88	-0.80	16.0	
	SM8: Overall, I think that the policy leads to goal 2[1,2]	1	5	3.88	0.99	-0.62	16.0	
	SM9: I think that the policy, in the long term, will lead to goal 3[1,2]	1	5	4.07	0.87	-0.79	16.7	
	SM10: I think that the policy, in the short term, will lead to goal 3[1,2]	1	5	4.17	0.76	-0.74	16.7	
	SM11: I think that the policy has already led to goal 3[1,2]	1	5	4.21	0.75	-0.77	16.6	
	SM12: Overall, I think that the policy leads to goal 3[1,2]	1	5	4.12	0.83	-0.81	16.6	
High client meaningfulness CM	CM1: With the policy I can better solve the problems of my clients[1]	1	5	4.26	0.76	-1.01	18.6	15.9
	CM1: In general, government policy enables me to better solve the problems of my clients [1]	*1*	*5*	*3.88*	*0.81*	*-0.54*	*0.5*	*22.5*
	CM2: The policy is contributing to the welfare of my clients[1]	1	5	4.18	0.79	-0.88	18.6	
	CM2: In general, government policy contributes to the welfare of my clients [1]	*1*	*5*	*3.58*	*0.94*	*-0.26*	*0.4*	
	CM3: I do not understand how my clients, in the short term, can profit from the policy	1	5	3.96	1.03	-1.03	18.3	
	CM3: In general, government policy enables me to help clients more efficiently[1]	*1*	*5*	*3.83*	*0.82*	*-0.49*	*0.5*	
	CM4: I think that the policy is ultimately favourable for my clients[1]	1	5	3.88	0.92	-0.46	18.2	
	CM4: Overall, I think government policy is ultimately favorable for my clients [1]	*1*	*5*	*3.46*	*0.98*	*-0.11*	*0.1*	

[1]Likert scale was reverse coded to ensure comparability.
[2]Question was asked concerning different goals: (1) transparency in costs and quality of mental healthcare, (2) efficiency and (3) patient choice.
Answer categories: 1=Completely disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Completely agree.
Calibration anchors:
Direct calibration: Values of 3 were recoded into set membership 0, values of 4 into 3, values of 5 into 4.
Conditions sets: Full nonmembership (4), crossover point (2.5), full membership (1).
W: Full nonmembership (1), crossover point (2.5), full membership (4).
Recoding method:
Conditions sets: 1= 1 (Fully powerful/meaningful), 2 = 0.8 (rather but not fully powerful/meaningful), 3 = 0.4 (rather not powerful/meaningful), 4 = 0.2 (rather but not fully not powerful/meaningful) 5 = 0 (fully not powerful/meaningful).
W: 5= 1 (Fully willing), 4 = 0.8 (rather but not fully willing), 3 = 0.4 (rather not willing), 2 = 0.2 (rather but not fully not willing) 1 = 0 (fully not willing).
No italics: dataset 1 (original N = 1317, final N = 1004), *italic*s: dataset 2 (original N = 1096, final N = 1087)

Table B2. Analyses of sufficiency, model selection and robustness tests for high implementation willingness, dataset 1, direct calibration

Frequency threshold	*Raw consistency threshold*	*Enhanced parsimonious solution term*	*Solution consistency*	*Solution PRI*	*Solution coverage*	*Z "Almost always sufficient" (0.8)*	*Z "Very often sufficient" (0.75)*	*Limited diversity*	*Model ambiguity*	*Robustness*	*Passes skewness test?*	*Model rank*
0.5 %	-	The first row has a PRI of 0.246						6.2 %				
(5)	0.922	SP*OP*SM*CM + SP*TP*SM*CM + TP*OP*SM*CM	0.912	0.588	0.471	0	3.7***		1	19	x	2
	0.910	**OP*SM*CM + SP*TP*SM*CM**	**0.907**	**0.600**	**0.488**	**-0.2**	**3.5*** **		**1**	**50**	**x**	**1**
1 %	0.906	OP*SM*CM + ~SP*TP*OP*SM	0.893	0.561	0.535	-0.5	3.2***	34.3 %	1	57	x	2
(10)	0.900	OP*SM*CM + ~SP*TP*OP*SM	0.893	0.561	0.535	-0.5	3.2***		1	57	x	2
	0.898	OP*SM*CM + ~SP*TP*OP*SM + (~TP*OP*CM)	0.870	0.509	0.559	-1.6	2.2**		2	43	x	3
5%	0.824	SP*TP*SM	0.882	0.596	0.525	-0.6	3.1***	75 %	2	43	x	4
(50)	0.802	SP*OP	0.766	0.484	0.66	-6.1	-1.9		1	29	x	4
	0.799	SP*OP + TP*OP	0.722	0.464	0.738	-11	-6.5		1	-	x	-

Untenable assumptions: ~TP*~OP + ~OP*~SP (contradicts statement of necessity). Directional expectations: SM → W, CM → W.
Rules for identifying raw consistency thresholds (truth table rows sorted by raw consistency): first threshold: above first row with PRI < 0.25, second threshold: above second row with PRI < 0.25, third threshold: above third row with PRI < 0.25. PRIs range from 0.172-0.573.
In case of model ambiguity, the table refers to the model with the highest consistency score.
*Significant at 0.1 level, **significant at 0.05 level, ***significant at 0.01 level.
See Table 1 for an explanation of the different parameters for model evaluation.
Bold: model chosen for interpretation.
Grey: Model does not pass consistency threshold 0.75.

Table B3. Analyses of sufficiency, model selection and robustness tests for high implementation willingness, dataset 1, recoding method

Frequency threshold	*Raw consistency threshold*	*Enhanced parsimonious solution term*	*Solution consistency*	*Solution PRI*	*Solution coverage*	*Z "Almost always sufficient" (0.8)*	*Limited diversity*	*Model ambiguity*	*Robustness*	*Passes skewness test?*	*Model rank*
0.5 % (5)	0.9468	~SP*OP*CM + ~TP*OP*CM + SP*TP*~OP*CM + SP*TP*SM*CM	0.925	0.793	0.489	9.6***	6.2 %	2	25	x	2
	0.9405	~SP*OP*CM + ~TP*OP*CM + SP*TP*~OP*CM + ~SP*TP*OP*SM + SP*TP*~OP*SM + SP*TP*SM*CM	0.907	0.753	0.548	6.9***		2	28	x	4
	0.9395	**OP*CM + SP*TP*CM + ~SP*TP*OP*SM + SP*TP*~OP*SM**	**0.890**	**0.727**	**0.565**	**6.3*** **		**1**	**42**	**x**	**1**
1 % (10)	0.9405	OP*SM*CM + ~SP*TP*OP*SM + ~TP*OP*CM	0.920	0.773	0.511	8.9***	34.3 %	2	30	x	4
	0.9395	OP*CM + ~SP*TP*OP*SM	0.900	0.738	0.534	7.8***		1	56	x	1
	0.928	OP*CM + OP*SM	0.866	0.716	0.626	4.2***		1	61	x	3
5% (50)	0.928	SP*TP*SM	0.917	0.786	0.498	9.3***	75 %	2	44	x	5
	0.858	SP*OP	0.820	0.642	0.702	-1.2		1	22	x	7
	0.839	SP*OP + TP*OP	0.780	0.601	0.765	-7		1	39	x	6

Untenable assumptions: ~TP*~OP + ~OP*~SP (contradicts statement of necessity). Directional expectations: SM → W, CM → W.
Rules for identifying raw consistency thresholds (truth table rows sorted by raw consistency): first threshold: above first row with PRI < 0.5, second threshold: above second row with PRI < 0.5, third threshold: above third row with PRI < 0.5. PRIs range from 0.239-0.873.
In case of model ambiguity, the table refers to the model with the highest consistency score.
*Significant at 0.1 level, **significant at 0.05 level, ***significant at 0.01 level.
See Table 1 for an explanation of the different parameters for model evaluation. Among the two first ranking models, we chose the one with higher coverage. The two models are making the same causal claims, but the chosen model explains additional cases.
Bold: model chosen for interpretation.

Table B4. Analyses of sufficiency, model selection and robustness tests for high implementation willingness, dataset 2, direct calibration

Frequency threshold	*Raw consistency threshold*	*Enhanced parsimonious solution term*	*Solution consistency*	*Solution PRI*	*Solution coverage*	*Z "Almost always sufficient" (0.8)*	*Limited diversity*	*Model ambiguity*	*Robustness*	*Passes skewness test?*	*Model rank*
0.5 %	0.937	SP*SM*CM + TP*OP*SM*CM	0.938	0.843	0.424	8.8***	3.1%	1	0.17	x	1
(5)	**0.934**	**OP*SM*CM + SP*SM*CM + TP*SM*CM**	**0.935**	**0.841**	**0.436**	**8.6*** **		**1**	**0.30**	**x**	**1**
	0.932	OP*SM*CM + TP*SM*CM + SP*~TP*~OP*CM	0.932	0.832	0.462	7.6***		1	0.26	x	2
1 %	0.929	TP*CM + (SP*CM)	0.918	0.812	0.485	7.3***	34.4%	2	0.61	x	4
(11)	0.926	SP*CM + ~SP*TP*SM	0.917	0.813	0.530	6.6***		1	0.5	x	3
	0.923	SP*CM + ~SP*TP*SM + TP*OP*SM + SP*~TP*~OP*SM	0.908	0.811	0.569	5.9***		1	0.33	x	5
5%	0.891	OP*CM	0.927	0.820	0.456	8***	78.1%	6	0.56	x	6
(54)	0.884	SP*TP*~OP + (OP*CM)	0.893	0.776	0.574	4.9***		6	0.39	x	7
	0.883	~SP*OP + SP*TP*~OP + (SP*CM)	0.854	0.741	0.673	1.8*		6	0.37	x	8

Untenable assumptions: ~SP*~OP*~TP (contradicts statement of necessity). Directional expectations: SM → W, CM → W.

Rules for identifying raw consistency thresholds (truth table rows sorted by raw consistency): first threshold: above first row with PRI < 0.65, second threshold: above second row with PRI < 0.65, third threshold: above third row with PRI < 0.65. PRIs range from 0.552-0.750.

In case of model ambiguity, the table refers to the model with the highest consistency score.

*Significant at 0.1 level, **significant at 0.05 level, ***significant at 0.01 level.

See Table 1 for an explanation of the different parameters for model evaluation. Among the two first ranking models, we chose the one with higher coverage. The two models are making the same causal claims, but the chosen model explains additional cases.

Bold: model chosen for interpretation.

Table B5. Analyses of sufficiency, model selection and robustness tests for low implementation willingness, dataset 1, direct calibration

Frequency threshold	*Raw consistency threshold*	*Enhanced parsimonious solution term*	*Solution consistency*	*Solution PRI*	*Solution coverage*	*Z "Almost always sufficient" (0.8)*	*Limited diversity*	*Model ambiguity*	*Robustness*	*Passes skewness test?*	*Model rank*
0.5 % (5)	0.972	~OP*~SM*CM + ~SP*~OP*CM + ~TP*~SM*CM + SP*~OP*SM*~CM	0.920	0.702	0.445	11.1***	6.2 %	1	11	x	4
	0.964	~SM*CM + ~OP*SM*~CM + (~SP*~OP*SM)	0.884	0.654	0.478	8.9***		2	48	x	6
	0.9613	**~SM*CM + ~OP*SM*~CM + ~SP*TP*~OP**	**0.879**	**0.693**	**0.538**	**7.9*** **		**1**	**52**	**x**	**1**
1 % (10)	0.9611	~SM*CM + ~OP*SM*~CM + ~SP*TP*~OP	0.879	0.693	0.538	7.9***	34.3 %	1	52	x	3
	0.9583	~SM*CM + ~OP*SM*~CM + TP*~OP*~CM	0.871	0.696	0.565	7***		1	52	x	2
	0.956	~SM*CM + ~OP*SM*~CM + TP*~OP*~CM + ~SP*TP*SM*~CM	0.861	0.678	0.573	6.2***		1	42	x	4
5% (50)	0.918	~OP*~SM + ~SP*TP*~SM + SP*~TP*~SM	0.850	0.735	0.744	2.4**	75 %	8	11	x	6
	0.901	~SP*~SM + ~TP*~SM	0.837	0.727	0.788	0.4		4	22	x	7
	0.899	~SM	0.783	0.672	0.881	-8.5		1	100	x	5

Untenable assumptions: OP*SM*CM + SP*TP*SM*CM (contradicts statement of sufficiency). Directional expectations: ~SP → ~W , ~TP → ~W , ~OP → ~W.
Rules for identifying raw consistency thresholds (truth table rows sorted by raw consistency): first threshold: above first row with PRI < 0.75, second threshold: above second row with PRI < 0.75, third threshold: above third row with PRI < 0.75. PRIs range from 0.427-0.888.
In case of model ambiguity, the table refers to the model with the highest consistency score.
*Significant at 0.1 level, **significant at 0.05 level, ***significant at 0.01 level.
See Table 1 for an explanation of the different parameters for model evaluation.
Bold: model chosen for interpretation.

Table B6. Analyses of sufficiency, model selection and robustness tests for low implementation willingness, dataset 1, recoding method

Frequency threshold	*Raw consistency threshold*	*Enhanced parsimonious solution term*	*Solution consistency*	*Solution PRI*	*Solution coverage*	*Z "Almost always sufficient" (0.8)*	*Limited diversity*	*Model ambiguity*	*Robustness*	*Passes skewness test?*	*Model rank*
0.5 % (5)	0.945	~TP*~OP*CM + ~SP*~OP*~SM*CM	0.909	0.484	0.337	8***	6.2 %	1	44	-	4
	0.943	~TP*~OP*CM + ~SP*~OP*~SM*CM + ~SP*TP*~OP*SM*~CM	0.899	0.490	0.373	6.9***		1	33	x	3
	0.939	~SP*~OP*CM + ~SP*~OP*SM + ~TP*~OP*CM + ~TP*~OP*SM	0.870	0.502	0.404	6***		1	38	x	2
1 % (10)	0.945	~TP*~OP*CM	0.915	0.470	0.313	9.3***	34.3 %	1	63	-	4
	0.934	~SP*TP*~OP + ~TP*~OP*CM + ~TP*~OP*SM	0.862	0.585	0.480	5.5***		1	38	x	2
	0.901	**~SP*~OP + ~TP*~OP + ~OP*~SM*~CM**	**0.818**	**0.627**	**0.640**	**1.8***		**1**	**46**	**x**	**1**
5% (50)	0.812	~TP*~CM + ~SP*~SM*~CM	0.762	0.573	0.773	-7.4	75 %	1	13	x	5
	0.794	~SM*~CM	0.712	0.523	0.817	-15		1	38	x	-
	-	- (there is no third row with PRI < 0.5)	-	-	-	-	-	-	-	-	-

Untenable assumptions: OP*CM + SP*TP*CM + ~SP*TP*OP*SM + SP*TP*~OP*SM (contradicts statement of sufficiency). Directional expectations: ~SP → ~W, ~TP → ~W, ~OP → ~W.

Rules for identifying raw consistency thresholds (truth table rows sorted by raw consistency): first threshold: above first row with PRI < 0.5, second threshold: above second row with PRI < 0.5, third threshold: above third row with PRI < 0.5. PRIs range from 0.111 - 0.749.

In case of model ambiguity, the table refers to the model with the highest consistency score.

*Significant at 0.1 level, **significant at 0.05 level, ***significant at 0.01 level.

See Table 1 for an explanation of the different parameters for model evaluation.

Bold: model chosen for interpretation.

Grey: Model does not pass consistency threshold 0.75.

Table B7. Analyses of sufficiency, model selection and robustness tests for low implementation willingness, dataset 2, direct calibration

Frequency threshold	*Raw consistency threshold*	*Enhanced parsimonious solution term*	*Solution consistency*	*Solution PRI*	*Solution coverage*	*Z "Almost always sufficient" (0.8)*	*Limited diversity*	*Model ambiguity*	*Robustness*	*Passes skewness test?*	*Model rank*
0.5 % (5)	0.882	~OP*~SM*CM + ~SP*~SM*CM + ~SP*~TP*~OP*CM + ~SP*~TP*OP*SM*~CM	0.835	0.358	0.555	2.8***	3.1%	1	50	x	2
	0.880	**~OP*~SM*CM + ~SP*~SM*CM + ~SP*~TP*~OP*CM + ~SP*~TP*OP*SM*~CM + SP*~TP*~OP*SM*~CM**	**0.826**	**0.360**	**0.567**	**2.2****		**1**	**50**	**x**	**1**
	0.8785	~OP*~SM*CM + ~SP*~SM*CM + ~SP*~TP*~OP*CM + ~SP*~TP*OP*SM*~CM + SP*~TP*~OP*SM*~CM	0.826	0.360	0.567	2.2**		1	50	x	1
1 % (11)	0.880	SP*~TP*~OP*SM*~CM	0.881	0.386	0.442	7.3***	34.4%	1	50	x	3
	0.868	~TP*SM*~CM + ~SP*~OP*SM*~CM	0.813	0.343	0.522	3.0***		1	67	x	4
	0.861	~TP*SM*~CM + ~SP*~OP*SM*~CM	0.813	0.343	0.522	3.0***		1	67	x	4
5% (54)	0.75	~SM	0.568	0.342	0.893	-27.4	78.1%	2		x	
	-	- (there is no second row with PRI < 0.35)	-	-	-	-		-	-	-	-
	-	- (there is no third row with PRI < 0.35)	-	-	-	-		-	-	-	-

Untenable assumptions: OP*SM*CM + SP*SM*CM + TP*SM*CM (contradicts statement of sufficiency). Directional expectations: ~SP → ~W, ~TP → ~W, ~OP → ~W.
Rules for identifying raw consistency thresholds (truth table rows sorted by raw consistency): first threshold: above first row with PRI < 0.35, second threshold: above second row with PRI < 0.35, third threshold: above third row with PRI < 0.35. PRIs range from 0.194 - 0.439.
In case of model ambiguity, the table refers to the model with the highest consistency score.
*Significant at 0.1 level, **significant at 0.05 level, ***significant at 0.01 level.
See Table 1 for an explanation of the different parameters for model evaluation.
Bold: model chosen for interpretation.
Grey: Model does not pass consistency threshold 0.75.

Table B8. Truth table for high implementation willingness 'W', dataset 1, direct calibration

Row Nr.	*SP*	*TP*	*OP*	*SM*	*CM*	*Outcome*	*N*	*Raw consistency*	*PRI*
28	1	1	0	1	1	1	5	0.924	0.246
16	0	1	1	1	1	1	9	0.924	0.324
32	1	1	1	1	1	1	40	0.923	0.573
24	1	0	1	1	1	1	7	0.922	0.272
12	0	1	0	1	1	0	5	0.922	0.220
8	0	0	1	1	1	1	11	0.922	0.332
22	1	0	1	0	1	0	5	0.909	0.183
15	0	1	1	1	0	0	15	0.906	0.309
14	0	1	1	0	1	0	8	0.905	0.235
11	0	1	0	1	0	0	10	0.905	0.236
27	1	1	0	1	0	0	9	0.902	0.202
6	0	0	1	0	1	0	11	0.899	0.201
30	1	1	1	0	1	0	12	0.897	0.238
18	1	0	0	0	1	0	10	0.894	0.164
23	1	0	1	1	0	0	22	0.894	0.299
26	1	1	0	0	1	0	9	0.893	0.177
10	0	1	0	0	1	0	8	0.892	0.144
31	1	1	1	1	0	0	56	0.890	0.481
19	1	0	0	1	0	0	16	0.885	0.187
7	0	0	1	1	0	0	35	0.884	0.340
2	0	0	0	0	1	0	14	0.879	0.112
3	0	0	0	1	0	0	28	0.866	0.197
21	1	0	1	0	0	0	54	0.823	0.225
29	1	1	1	0	0	0	89	0.822	0.360
25	1	1	0	0	0	0	47	0.817	0.184
13	0	1	1	0	0	0	75	0.801	0.238
9	0	1	0	0	0	0	53	0.798	0.161
17	1	0	0	0	0	0	68	0.790	0.180
5	0	0	1	0	0	0	121	0.761	0.254
1	0	0	0	0	0	0	146	0.702	0.172
4	0	0	0	1	1	0	2	0.920	0.172
20	1	0	0	1	1	0	4	0.919	0.196

Raw consistency threshold: 0.910. Frequency threshold: 5.

Parsimonious solution (without exclusion of contradictory rows and untenable assumptions): SM*CM → W (solution consistency 0.903, solution coverage 0.500).

Contradictory rows and untenable assumptions: ~TP*~OP + ~OP*~SP → W (contradicts statement of necessity).

Conservative solution: OP*SM*CM + SP*TP*SM*CM → W (solution consistency 0.907, solution coverage 0.488).

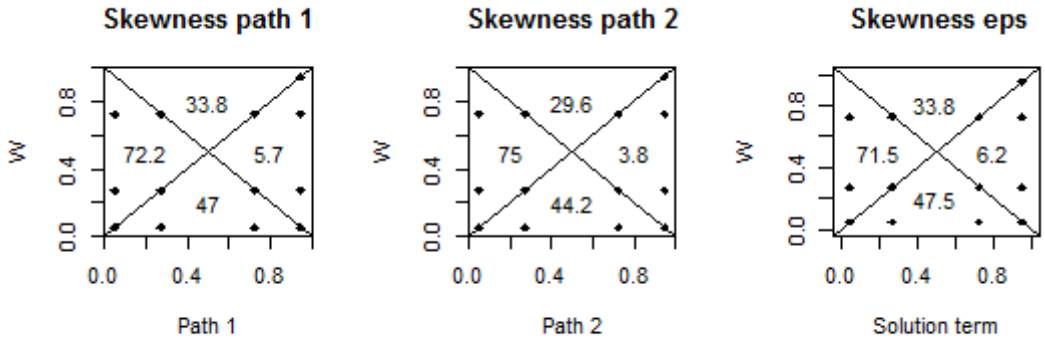
Enhanced parsimonious solution (under exclusion of contradictory rows and untenable assumptions) OP*SM*CM + SP*TP*SM*CM → W (solution consistency 0.907, solution coverage 0.488).
Directional expectations: SM → W, CM → W.

Note: the untenable assumptions applied under ESA combine the two necessary conditions SP+OP and TP+OP with the logical AND. However, the condition (SP+OP)*(TP+OP) in fact has a consistency score necessity of 0.870 for W (Coverage 0.672, RoN 0.645). This is because, while according to formal logic, it always holds that if A ← Y and B ←Y, then also A*B ← Y, this may no longer hold when encountering less-than-perfect subset relations (here: quasi-necessity), as is usually the case with empirical data.

Table B9. Simplifying assumptions for analysis of W, dataset 1, direct calibration

SP	*TP*	*OP*	*SM*	*CM*	*Easy counterfactual (used for intermediate solution) (N = 0)?*
0	0	0	1	1	-
1	0	0	1	1	-

Figure B1. Skewness test, intermediate solution for outcome 'W', dataset 1, direct calibration



Per cent of cases falling within intersecting area of the XY-plot with two diagonals (Schneider and Wagemann 2012).

Table B10. Truth table for high implementation willingness 'W', dataset 1, recoding method

Row Nr.	*SP*	*TP*	*OP*	*SM*	*CM*	*Outcome*	*N*	*Raw consistency*	*PRI*
28	1	1	0	1	1	1	5	0.969	0.726
12	0	1	0	1	1	0	5	0.967	0.635
16	0	1	1	1	1	1	9	0.966	0.776
8	0	0	1	1	1	1	11	0.964	0.738
32	1	1	1	1	1	1	40	0.962	0.873
24	1	0	1	1	1	1	7	0.959	0.740
14	0	1	1	0	1	1	8	0.953	0.629
6	0	0	1	0	1	1	11	0.951	0.578
22	1	0	1	0	1	1	5	0.950	0.579
26	1	1	0	0	1	1	9	0.947	0.534
11	0	1	0	1	0	0	10	0.947	0.512
10	0	1	0	0	1	0	8	0.947	0.443
15	0	1	1	1	0	1	15	0.944	0.662
27	1	1	0	1	0	1	9	0.941	0.534
18	1	0	0	0	1	0	10	0.940	0.450
30	1	1	1	0	1	1	12	0.940	0.648
2	0	0	0	0	1	0	14	0.939	0.380
7	0	0	1	1	0	0	35	0.931	0.613
31	1	1	1	1	0	0	56	0.930	0.742
23	1	0	1	1	0	0	22	0.930	0.630
19	1	0	0	1	0	0	16	0.927	0.441
3	0	0	0	1	0	0	28	0.920	0.415
21	1	0	1	0	0	0	54	0.868	0.468
25	1	1	0	0	0	0	47	0.867	0.384
29	1	1	1	0	0	0	89	0.866	0.563
13	0	1	1	0	0	0	75	0.857	0.439
17	1	0	0	0	0	0	68	0.838	0.320
9	0	1	0	0	0	0	53	0.836	0.277
5	0	0	1	0	0	0	121	0.823	0.387
1	0	0	0	0	0	0	146	0.766	0.239

Raw consistency threshold: 0.9395 . Frequency threshold: 5.

Parsimonious solution (without exclusion of contradictory rows and untenable assumptions): OP*CM + SP*CM + TP*CM + ~SP*TP*SM + TP*~OP*SM → W (solution consistency 0.873, solution coverage 0.586).

Contradictory rows and untenable ssumptions: ~TP*~OP + ~OP*~SP → W (contradicts statement of necessity).

Conservative solution: OP*CM + SP*TP*CM + ~SP*TP*OP*SM + SP*TP*~OP*SM → W (solution consistency 0.890, solution coverage 0.565).

Enhanced parsimonious solution (under exclusion of contradictory rows and untenable assumptions) OP*CM + SP*TP*CM + ~SP*TP*OP*SM + SP*TP*~OP*SM → W (solution consistency 0.890, solution coverage 0.565). Directional expectations: SM → W, CM → W.

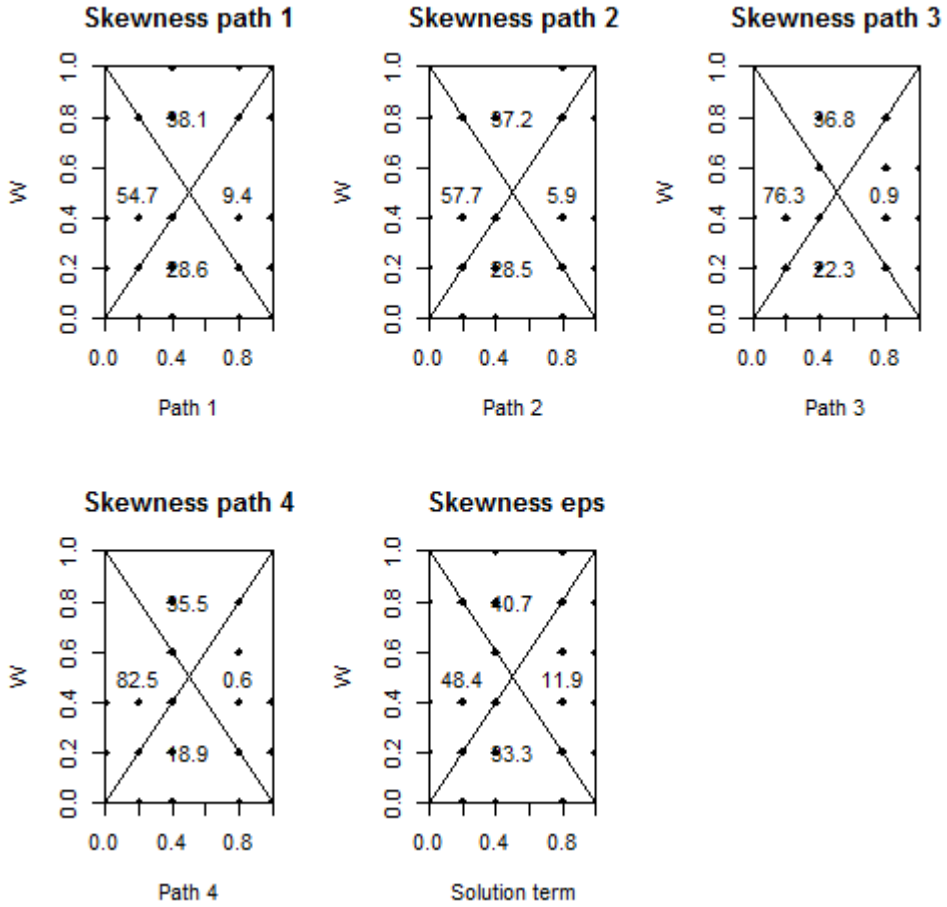
Note: the untenable assumptions applied under ESA combine the two necessary conditions SP+OP and TP+OP with the logical AND. However, the condition (SP+OP)*(TP+OP) in fact has a consistency score necessity of

0.887 for W (Coverage 0.707, RoN 0.646). This is because, while according to formal logic, it always holds that if A ← Y and B ←Y, then also A*B ← Y, this may no longer hold when encountering less-than-perfect subset relations (here: quasi-necessity), as is usually the case with empirical data.

Table B11. Simplifying assumptions for analysis of W, dataset 1, recoding method

SP	*TP*	*OP*	*SM*	*CM*	*Easy counterfactual (used for intermediate solution) (N = 0)?*
1	0	0	1	1	-

Figure B2. Skewness test, intermediate solution for outcome 'W', dataset 1, recoding method



Per cent of cases falling within intersecting area of the XY-plot with two diagonals (Schneider and Wagemann 2012).

Table B12. Truth table for high implementation willingness 'W', dataset 2, direct calibration

Row Nr.	*SP*	*TP*	*OP*	*SM*	*CM*	*Outcome*	*N*	*Raw consistency*	*PRI*
16	0	1	1	1	1	1	32	0.943	0.750
28	1	1	0	1	1	1	16	0.939	0.694
24	1	0	1	1	1	1	11	0.938	0.681
20	1	0	0	1	1	1	9	0.937	0.661
32	1	1	1	1	1	1	97	0.937	0.806
4	0	0	0	1	1	0	7	0.937	0.642
8	0	0	1	1	1	1	7	0.937	0.652
12	0	1	0	1	1	1	12	0.935	0.653
18	1	0	0	0	1	0	5	0.933	0.631
22	1	0	1	0	1	0	9	0.932	0.647
30	1	1	1	0	1	0	17	0.930	0.672
7	0	0	1	1	0	0	8	0.930	0.626
14	0	1	1	0	1	0	5	0.929	0.617
11	0	1	0	1	0	0	16	0.929	0.647
6	0	0	1	0	1	0	5	0.929	0.610
15	0	1	1	1	0	0	20	0.928	0.663
19	1	0	0	1	0	0	11	0.925	0.614
31	1	1	1	1	0	0	52	0.924	0.711
3	0	0	0	1	0	0	18	0.922	0.622
26	1	1	0	0	1	0	6	0.922	0.586
27	1	1	0	1	0	0	33	0.922	0.662
10	0	1	0	0	1	0	6	0.920	0.561
23	1	0	1	1	0	0	17	0.918	0.604
21	1	0	1	0	0	0	34	0.910	0.623
25	1	1	0	0	0	0	75	0.890	0.620
5	0	0	1	0	0	0	48	0.888	0.566
13	0	1	1	0	0	0	103	0.883	0.638
17	1	0	0	0	0	0	73	0.883	0.593
9	0	1	0	0	0	0	83	0.877	0.594
29	1	1	1	0	0	0	148	0.874	0.650
1	0	0	0	0	0	0	103	0.852	0.552

Raw consistency threshold: 0.934. Frequency threshold: 5.

Parsimonious solution (without exclusion of contradictory rows and untenable assumptions):
SM*CM → W (solution consistency 0.934, solution coverage 0.441).

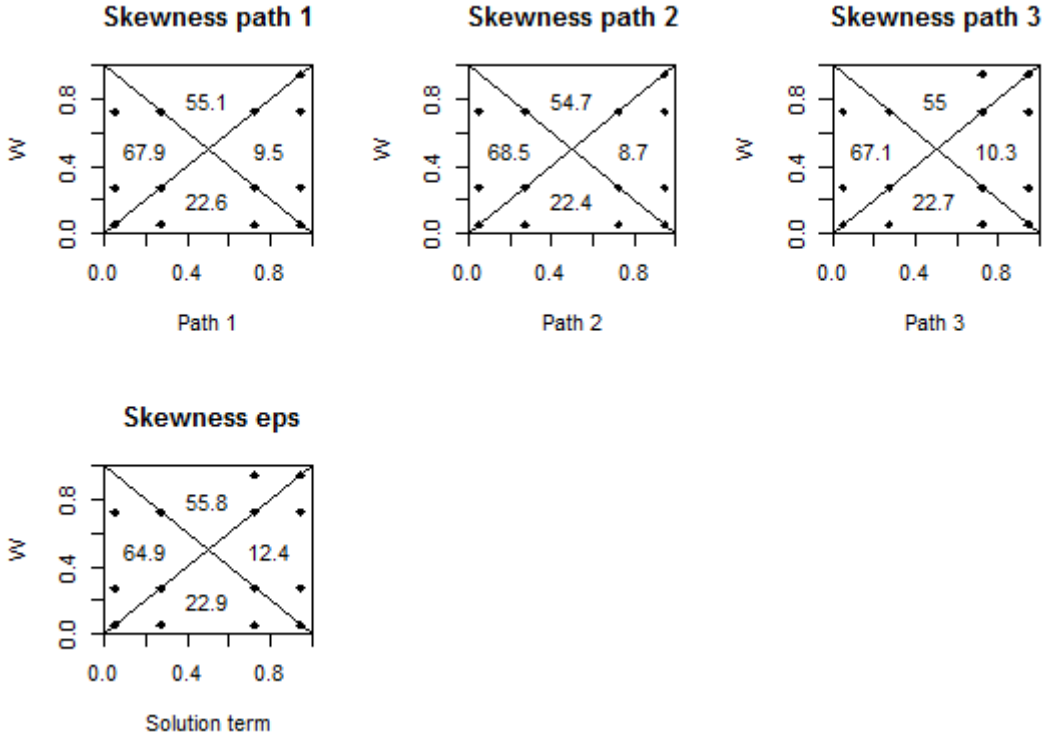
Contradictory rows and untenable assumptions: ~SP*~TP*~OP → W (contradicts statement of necessity).

Conservative solution: OP*SM*CM + SP*SM*CM + TP*SM*CM→ W (solution consistency 0.935, solution coverage 0.436).

Enhanced parsimonious solution (under exclusion of contradictory rows and untenable assumptions) OP*SM*CM + SP*SM*CM + TP*SM*CM → W (solution consistency 0.935, solution coverage 0.436).

Directional expectations: SM → W, CM → W. No simplifying assumptions applied.

Figure B3. Skewness test, intermediate solution for outcome 'W', dataset 2, direct calibration



Per cent of cases falling within intersecting area of the XY-plot with two diagonals (Schneider and Wagemann 2012).

Table B13. Truth table for high implementation willingness 'W', dataset 2, recoding method

Row Nr.	*SP*	*TP*	*OP*	*SM*	*CM*	*Outcome*	*N*	*Raw consistency*	*PRI*
16	0	1	1	1	1	0	32	0.996	0.986
8	0	0	1	1	1	0	7	0.995	0.975
4	0	0	0	1	1	0	7	0.995	0.971
20	1	0	0	1	1	0	9	0.994	0.972
30	1	1	1	0	1	0	17	0.993	0.977
12	0	1	0	1	1	0	12	0.993	0.966
14	0	1	1	0	1	0	5	0.993	0.966
24	1	0	1	1	1	0	11	0.992	0.969
11	0	1	0	1	0	0	16	0.992	0.964
28	1	1	0	1	1	0	16	0.992	0.970
7	0	0	1	1	0	0	8	0.992	0.958
18	1	0	0	0	1	0	5	0.992	0.955
6	0	0	1	0	1	0	5	0.991	0.954
22	1	0	1	0	1	0	9	0.990	0.955
26	1	1	0	0	1	0	6	0.989	0.949
10	0	1	0	0	1	0	6	0.989	0.940
3	0	0	0	1	0	0	18	0.989	0.944
19	1	0	0	1	0	0	11	0.988	0.945
15	0	1	1	1	0	0	20	0.988	0.955
27	1	1	0	1	0	0	33	0.985	0.943
32	1	1	1	1	1	0	97	0.984	0.965
23	1	0	1	1	0	0	17	0.983	0.929
31	1	1	1	1	0	0	52	0.981	0.947
21	1	0	1	0	0	0	34	0.971	0.889
25	1	1	0	0	0	0	75	0.960	0.872
5	0	0	1	0	0	0	48	0.959	0.836
13	0	1	1	0	0	0	103	0.953	0.859
17	1	0	0	0	0	0	73	0.951	0.829
9	0	1	0	0	0	0	83	0.951	0.834
29	1	1	1	0	0	0	148	0.935	0.847
1	0	0	0	0	0	0	103	0.929	0.763

Note: Overdetermined truth table: The raw consistency and PRI values are generally too high (range from 0.763 to 0.986) to allow for setting a meaningful raw consistency threshold. This indicates that all truth table rows are sufficient for W. See also Table B20. Therefore, no analysis of sufficiency for the outcome W can be performed with dataset 2, using the recoding method.

Table B14. Truth table for low implementation willingness '~W', dataset 1, direct calibration

Row Nr.	*SP*	*TP*	*OP*	*SM*	*CM*	*Outcome*	*N*	*Raw consistency*	*PRI*
2	0	0	0	0	1	1	14	0.985	0.888
10	0	1	0	0	1	1	8	0.982	0.856
22	1	0	1	0	1	1	5	0.980	0.817
18	1	0	0	0	1	1	10	0.979	0.836
12	0	1	0	1	1	1	5	0.978	0.780
26	1	1	0	0	1	1	9	0.977	0.823
27	1	1	0	1	0	1	9	0.975	0.798
28	1	1	0	1	1	0	5	0.975	0.754
6	0	0	1	0	1	1	11	0.975	0.799
19	1	0	0	1	0	1	16	0.974	0.813
24	1	0	1	1	1	0	7	0.971	0.728
14	0	1	1	0	1	1	8	0.971	0.765
11	0	1	0	1	0	1	10	0.971	0.764
30	1	1	1	0	1	1	12	0.968	0.762
3	0	0	0	1	0	1	28	0.967	0.803
16	0	1	1	1	1	0	9	0.963	0.676
9	0	1	0	0	0	1	53	0.961	0.839
8	0	0	1	1	1	0	11	0.961	0.668
25	1	1	0	0	0	0	47	0.959	0.816
15	0	1	1	1	0	0	15	0.958	0.691
23	1	0	1	1	0	0	22	0.955	0.701
17	1	0	0	0	0	0	68	0.954	0.820
21	1	0	1	0	0	0	54	0.949	0.775
7	0	0	1	1	0	0	35	0.940	0.660
1	0	0	0	0	0	0	146	0.938	0.828
13	0	1	1	0	0	0	75	0.938	0.762
5	0	0	1	0	0	0	121	0.917	0.741
29	1	1	1	0	0	0	89	0.900	0.640
31	1	1	1	1	0	0	56	0.898	0.519
32	1	1	1	1	1	0	40	0.897	0.427

Raw consistency threshold: 0.879. Frequency threshold: 5.

Parsimonious solution (without exclusion of contradictory rows and untenable assumptions): ~OP*SM + ~SM*CM + ~SP*TP*CM + SP*~TP*CM + ~SP*TP*~OP → ~W (solution consistency 0.856, solution coverage 0.543).

Contradictory rows and untenable assumptions: OP*SM*CM + SP*TP*SM*CM → ~W (contradicts statement of sufficiency).

Conservative solution: ~SM*CM + ~OP*SM*~CM + ~SP*TP*~OP → ~W (solution consistency 0.879, solution coverage 0.538).

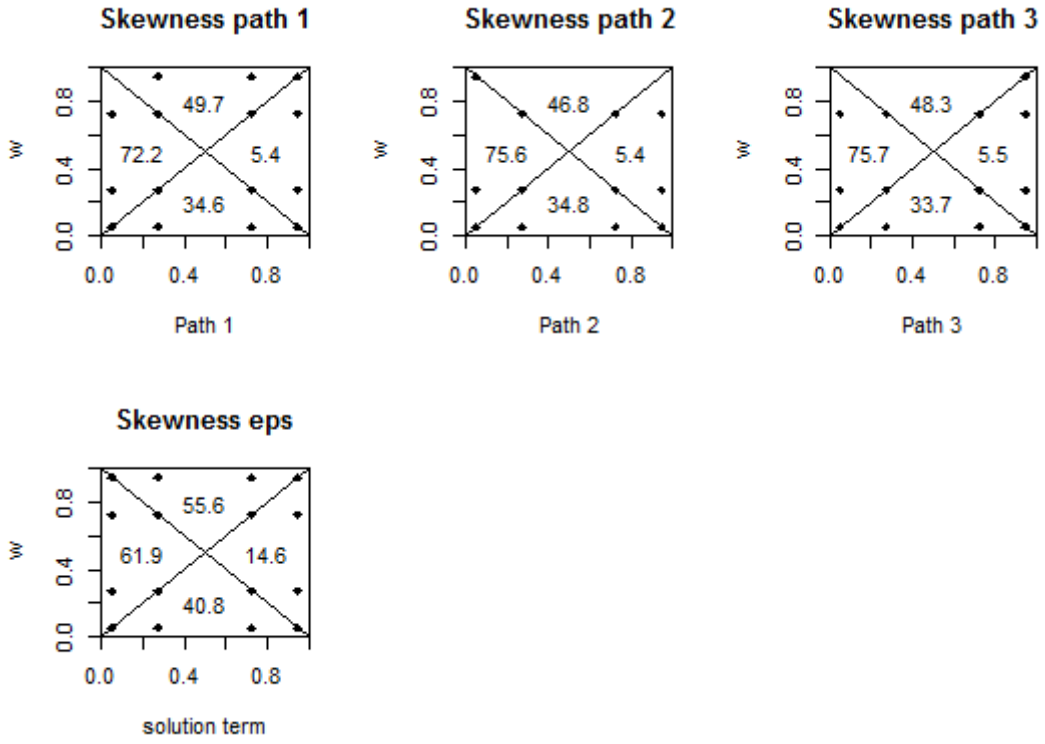
Enhanced parsimonious solution (under exclusion of contradictory rows and untenable assumptions) ~SM*CM + ~OP*SM*~CM + ~SP*TP*~OP → ~W (solution consistency 0.879, solution coverage 0.538).

Directional expectations: ~SP → ~W, ~TP → ~W, ~OP → ~W.

Table B15. Simplifying assumptions for analysis of ~W, dataset 1, direct calibration

SP	*TP*	*OP*	*SM*	*CM*	*Easy counterfactual (used for intermediate solution) (N = 0)?*
0	0	0	1	1	-
1	0	0	1	1	-

Figure B4. Skewness test, intermediate solution for outcome '~W', dataset 1, direct calibration



Per cent of cases falling within intersecting area of the XY-plot with two diagonals (Schneider and Wagemann 2012).

Table B16. Truth table for low implementation willingness '~W', dataset 1, recoding method

Row Nr.	*SP*	*TP*	*OP*	*SM*	*CM*	*Outcome*	*N*	*Raw consistency*	*PRI*
2	0	0	0	0	1	1	14	0.961	0.606
18	1	0	0	0	1	1	10	0.950	0.537
11	0	1	0	1	0	1	10	0.944	0.488
3	0	0	0	1	0	1	28	0.942	0.575
19	1	0	0	1	0	1	16	0.941	0.549
9	0	1	0	0	0	1	53	0.934	0.710
6	0	0	1	0	1	0	11	0.933	0.422
1	0	0	0	0	0	1	146	0.923	0.749
17	1	0	0	0	0	1	68	0.916	0.647
25	1	1	0	0	0	1	47	0.914	0.603
8	0	0	1	1	1	0	11	0.900	0.262
7	0	0	1	1	0	0	35	0.891	0.387
30	1	1	1	0	1	0	12	0.889	0.346
13	0	1	1	0	0	0	75	0.887	0.557
15	0	1	1	1	0	0	15	0.887	0.325
5	0	0	1	0	0	0	121	0.882	0.592
21	1	0	1	0	0	0	54	0.881	0.522
23	1	0	1	1	0	0	22	0.879	0.364
29	1	1	1	0	0	0	89	0.811	0.385
31	1	1	1	1	0	0	56	0.793	0.238
32	1	1	1	1	1	0	40	0.738	0.111

Raw consistency threshold: 0.901. Frequency threshold: 10.

Parsimonious solution (without exclusion of contradictory rows and untenable assumptions): ~OP + ~TP*~SM*CM→ w (solution consistency 0.786, solution coverage 0.689.

Contradictory rows and untenable assumptions: OP*CM + SP*TP*CM + ~SP*TP*OP*SM + SP*TP*~OP*SM→ ~W (contradicts statement of sufficiency).

Conservative solution: ~OP*~SM*~CM + ~SP*~OP*~CM + ~TP*~OP*~CM + ~TP*~OP*~SM → ~W (solution consistency 0.831, solution coverage 0.632).

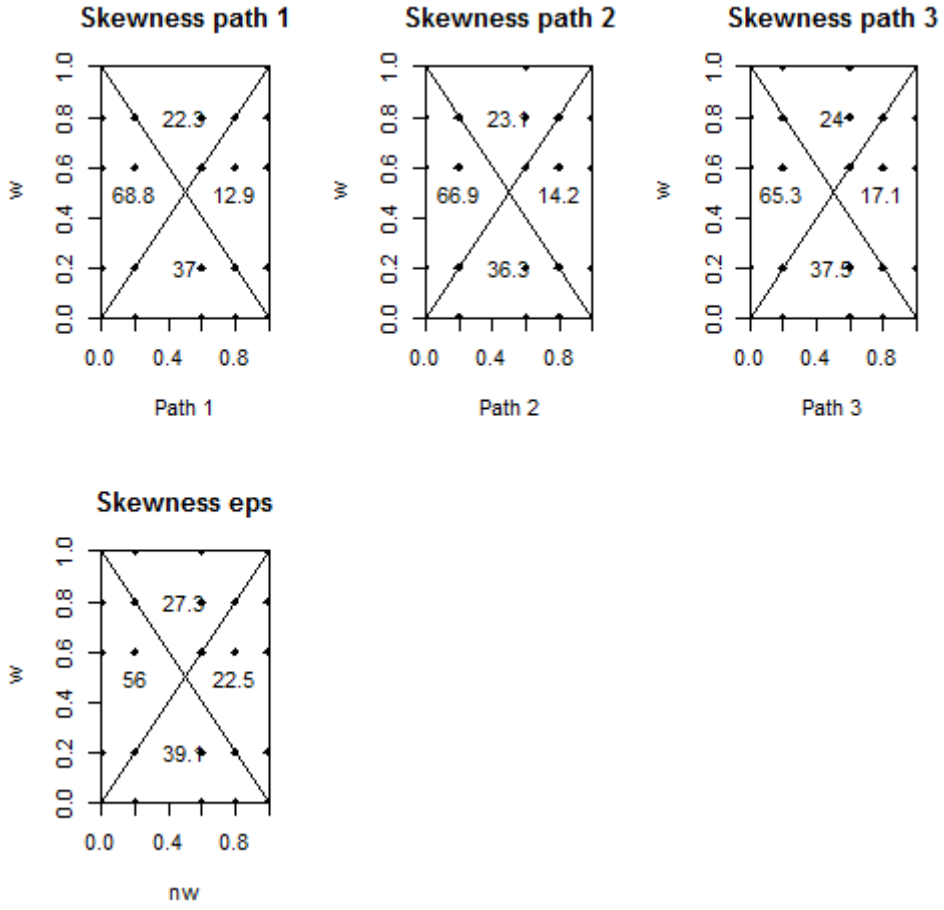
Enhanced parsimonious solution (under exclusion of contradictory rows and untenable assumptions) ~SP*~OP + ~TP*~OP + ~OP*~SM*~CM → ~W (solution consistency 0.818, solution coverage 0.640).

Directional expectations: ~SP → ~W , ~TP → ~W , ~OP → ~W.

Table B17. Simplifying assumptions for analysis of ~W, dataset 1, recoding method

SP	*TP*	*OP*	*SM*	*CM*	*Easy counterfactual (used for intermediate solution) (N = 4)?*
0	0	0	1	1	X
0	1	0	0	1	X
0	1	0	1	1	X
1	0	0	1	1	X
1	0	1	0	1	-
1	1	0	0	1	-
1	1	0	1	0	-
1	1	0	1	1	-

Figure B5. Skewness test, intermediate solution for outcome '~W', dataset 1, recoding method



Percent of cases falling within intersecting area of the XY-plot with two diagonals (Schneider and Wagemann 2012).

Table B18. Truth table for low implementation willingness '~W', dataset 2, direct calibration

Row Nr.	*SP*	*TP*	*OP*	*SM*	*CM*	*Outcome*	*N*	*Raw consistency*	*PRI*
10	0	1	0	0	1	1	6	0.898	0.439
26	1	1	0	0	1	1	6	0.890	0.414
6	0	0	1	0	1	1	5	0.888	0.390
4	0	0	0	1	1	1	7	0.886	0.358
14	0	1	1	0	1	1	5	0.885	0.383
18	1	0	0	0	1	1	5	0.885	0.369
7	0	0	1	1	0	1	8	0.883	0.374
8	0	0	1	1	1	0	7	0.881	0.348
19	1	0	0	1	0	1	11	0.881	0.386
12	0	1	0	1	1	0	12	0.879	0.347
20	1	0	0	1	1	0	9	0.878	0.339
23	1	0	1	1	0	0	17	0.875	0.396
22	1	0	1	0	1	0	9	0.875	0.353
3	0	0	0	1	0	0	18	0.872	0.378
11	0	1	0	1	0	0	16	0.869	0.353
24	1	0	1	1	1	0	11	0.868	0.319
28	1	1	0	1	1	0	16	0.861	0.306
15	0	1	1	1	0	0	20	0.859	0.337
30	1	1	1	0	1	0	17	0.857	0.328
5	0	0	1	0	0	0	48	0.854	0.434
21	1	0	1	0	0	0	34	0.851	0.373
27	1	1	0	1	0	0	33	0.847	0.338
17	1	0	0	0	0	0	73	0.830	0.407
16	0	1	1	1	1	0	32	0.828	0.250
25	1	1	0	0	0	0	75	0.821	0.380
9	0	1	0	0	0	0	83	0.820	0.406
1	0	0	0	0	0	0	103	0.817	0.448
31	1	1	1	1	0	0	52	0.813	0.289
13	0	1	1	0	0	0	103	0.794	0.362
29	1	1	1	0	0	0	148	0.765	0.350
32	1	1	1	1	1	0	97	0.739	0.194

Raw consistency threshold: 5. Frequency threshold: 0.880.

Parsimonious solution (without exclusion of contradictory rows and untenable assumptions): ~OP*~SM*CM + ~SP*~SM*CM + ~SP*~TP*CM + ~SP*~TP*OP*SM + SP*~TP*~OP*SM*~CM → ~W (solution consistency 0.818, solution coverage 0.568).

Contradictory rows and untenable assumptions: OP*SM*CM + SP*SM*CM + TP*SM*CM → ~W (contradicts statement of sufficiency).

Conservative solution (solution consistency 0.828, solution coverage 0.567).

M1: ~SP*OP*~SM*CM + SP*~OP*~SM*CM + ~SP*~TP*~OP*SM*CM + ~SP*~TP*OP*SM*~CM + SP*~TP*~OP*SM*~CM + (~SP*TP*~SM*CM) → ~W

M2: ~SP*OP*~SM*CM + SP*~OP*~SM*CM + ~SP*~TP*~OP*SM*CM + ~SP*~TP*OP*SM*~CM +

SP*~TP*~OP*SM*~CM + (TP*~OP*~SM*CM) → ~W

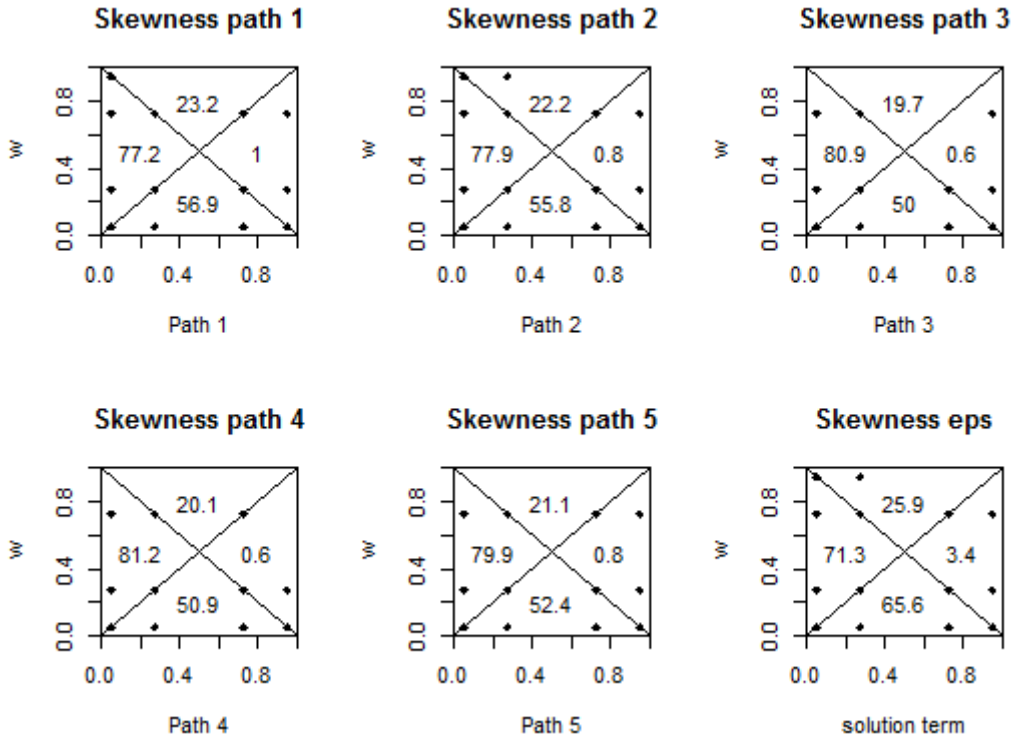
Enhanced parsimonious solution (under exclusion of contradictory rows and untenable assumptions) ~OP*~SM*CM + ~SP*~SM*CM + ~SP*~TP*~OP*CM + ~SP*~TP*OP*SM*~CM + SP*~TP*~OP*SM*~CM → ~W (solution consistency 0.826, solution coverage 0.567).

Directional expectations: ~SP → ~W, ~TP → ~W, ~OP → ~W.

Table B19. Simplifying assumptions for analysis of ~W, dataset 2, direct calibration

SP	*TP*	*OP*	*SM*	*CM*	*Easy counterfactual (used for intermediate solution) (N = 1)?*
0	0	0	0	1	X

Figure B6. Skewness test, intermediate solution for outcome '~W', dataset 2, direct calibration



Per cent of cases falling within intersecting area of the XY-plot with two diagonals (Schneider and Wagemann 2012).

Table B20. Truth table for low implementation willingness '~W', dataset 2, recoding method

Row Nr.	*SP*	*TP*	*OP*	*SM*	*CM*	*Outcome*	*N*	*Raw consistency*	*PRI*
10	0	1	0	0	1	0	6	0.830	0.060
18	1	0	0	0	1	0	5	0.820	0.045
6	0	0	1	0	1	0	5	0.819	0.046
4	0	0	0	1	1	0	7	0.819	0.029
3	0	0	0	1	0	0	18	0.807	0.056
8	0	0	1	1	1	0	7	0.806	0.025
7	0	0	1	1	0	0	8	0.806	0.036
26	1	1	0	0	1	0	6	0.803	0.051
12	0	1	0	1	1	0	12	0.801	0.034
19	1	0	0	1	0	0	11	0.799	0.049
11	0	1	0	1	0	0	16	0.792	0.036
22	1	0	1	0	1	0	9	0.791	0.045
20	1	0	0	1	1	0	9	0.790	0.028
14	0	1	1	0	1	0	5	0.789	0.034
5	0	0	1	0	0	0	48	0.788	0.152
23	1	0	1	1	0	0	17	0.780	0.067
1	0	0	0	0	0	0	103	0.770	0.233
24	1	0	1	1	1	0	11	0.763	0.027
17	1	0	0	0	0	0	73	0.761	0.168
21	1	0	1	0	0	0	34	0.761	0.100
15	0	1	1	1	0	0	20	0.750	0.041
9	0	1	0	0	0	0	83	0.749	0.154
27	1	1	0	1	0	0	33	0.747	0.057
28	1	1	0	1	1	0	16	0.747	0.030
25	1	1	0	0	0	0	75	0.725	0.118
16	0	1	1	1	1	0	32	0.719	0.014
30	1	1	1	0	1	0	17	0.713	0.023
13	0	1	1	0	0	0	103	0.708	0.129
31	1	1	1	1	0	0	52	0.657	0.051
29	1	1	1	0	0	0	148	0.631	0.131
32	1	1	1	1	1	0	97	0.551	0.027

Note: The raw consistency values are acceptable, but the PRI values are generally too low (range from 0.027 to 0.060) to allow for setting a meaningful raw consistency threshold. The PRI values indicate that these truth table rows are equally, or more, sufficient for the outcome W. There is no configuration that is more clearly sufficient for w than for W. See also Table B13. Therefore, no analysis of sufficiency for w can be performed with dataset 2, using the recoding method..

Table B21. Sufficient conditions for implementation willingness (intermediate solution), dataset 1, direct calibration

	Hypothesis 2: *Low implementation willingness (~W)*			**Hypothesis 3:** *High implementation willingness (W)*	
	Dataset 1[2]			*Dataset 1*[1]	
Configuration	**1**	**2**	**3**	**1**	**2**
High strategic powerfulness SP			○		●
High tactical powerfulness TP			●		●
High operational powerfulness OP		○	○	●	
High societal meaningfulness SM	○	●		●	●
High client meaningfulness CM	●	○		●	●
Consistency	0.927***	0.929***	0.936***	0.909***	0.921***
Raw coverage	0.404	0.386	0.420	0.481	0.431
Unique coverage	0.059	0.033	0.061	0.057	0.007
Solution consistency			0.879***		0.907***
Solution PRI			0.693		0.600
Solution coverage			0.538		0.488

Black circles indicate the presence of a condition, and white circles its absence. Blank spaces indicate the irrelevance of a condition. Raw consistency and frequency thresholds see Tables B8 and B14, online appendix. *Significant at 0.1 level, **significant at 0.05 level, ***significant at 0.01 level. [1]"very often sufficient" [2]"almost always sufficient.

Table B22. Evaluation of hypothesis 2, dataset 1, direct calibration

		Empirics	
		Detected in solution	Not detected in solution
Theory	Hypothesized	**~SP*~TP*~OP***~SM*CM + **~SP*~TP*~OP***SM*~CM 3 % (~W): support theory 1.2 % (W): contradict theory & solution	**~SP*~TP*~OP***~SM*~CM + **~SP*~TP*~OP***SM*CM 11.2 % (~W): support theory 3.6 % (W): delimit theory
	Not hypothesized	OP*~SM*CM + SP*~SM*CM + SP*~OP*SM*~CM + TP*~SM*CM + TP*~OP*SM*~CM + ~SP*TP*~OP 8.8 % (~W): extend theory 6.8 % (W): empirical contradictions	OP*~CM + SP*OP*~SM*~CM + ~TP*OP*~SM*~CM + SP*OP*SM*CM + ~TP*OP*SM*CM + OP*SM + SP*OP*~CM + SP*~SM*~CM + SP*~TP*~SM*~CM + SP*SM*CM + SP*~TP*SM*CM + SP*OP*SM + TP*OP*~CM + SP*TP*~SM*~CM + SP*TP*SM*CM + TP*OP*SM 33.8 % (~W): point to overlooked explanations 31.8 % % (W): support theory
Supports theory		Extends theory	Delimits theory

Based on Schneider and Wagemann (2012: 301).
Bold: hypothesized combinations.
Hypothesis 2: ~SP*~TP*~OP → ~W.

Table B23. Evaluation of hypothesis 3, dataset 1, direct calibration

		Empirics	
		Detected in solution	*Not detected in solution*
Theory	*Hypothesized*	**OP*SM*CM** + **SP*TP*OP*SM*CM** + **SP*OP*SM*CM** + **SP*TP*SM*CM** + **TP*OP*SM*CM** 5.9 % (W): support theory 1.3 % (~W): contradict theory & solution	**OP*SM***~CM + **OP***~SM***CM** + **SP*SM***~CM + **SP***~TP*~OP***SM** + **SP***~TP*~OP***CM** + **SP***~SM***CM** + **TP*SM***~CM + ~SP***TP***~OP***SM** + ~SP***TP***~OP***CM** + **TP***~SM***CM** 14 % (W): support theory 9.4 % (~W): delimit theory
	Not hypothesized	Empty set	~SP*~TP*~OP*~CM + ~SP*~TP*~OP + ~SP*~TP*~OP*~SM + ~SM*~CM + ~SP*~OP*~SM*~CM + ~TP*~OP*~SM*~CM 23.4 % (W): point to overlooked explanations 46 % (~W): support theory
Supports theory		Extends theory	Delimits theory

Based on Schneider and Wagemann (2012: 301).
Bold: hypothesized combinations.
Hypothesis 3: OP*SM + OP*CM + SP*SM + SP*CM + TP*SM + TP*CM → W.

Online Appendix II: Regression results

Box B1. Results of simple regression analysis and polynomial regression analysis

We conducted two alternative analyses to illustrate how our results differ from those obtained with methods that only analyze symmetric relationships.

Alternative analysis 1: Simple regression analysis

Interpreted in symmetric terms, interpretation 1 would simply state that the more a frontline worker feels powerful, the more he or she is willing to implement a policy (H1 & H2). OLS regression (using SPSS, version 23) allows us to evaluate the relationship between powerfulness (strategic, tactical and operational) and implementation willingness in both dataset 1 (healthcare) and dataset 2 (education). The results are shown in table 1.

The regression results ('model 1' in table 1) indicate that powerfulness positively correlates with implementation willingness. If frontline workers more strongly experience powerfulness they tend to have a higher implementation willingness and, vice versa, if they less strongly experience powerfulness, they tend to have lower implementation willingness. Please note that in both dataset 1 and 2 these results are statistically significant, with the exception of tactical powerfulness and implementation willingness in study 2 (β=0.04, n.s.). These findings are compatible with the findings from large-N set-theoretic configurational analysis: they both indicate that powerfulness and implementation willingness are positively related. However, results using this method do not provide information on the asymmetric effect of powerfulness for high implementation willingness.

In QCA, the notion of controls is usually not part of the analysis, as QCA does not consider isolating and estimating independent effects of causal variables as the central goal of analysis, but instead focuses on combinations of causally relevant conditions (Fiss et al. 2013). In contrast, incorporating solutions into a regression analysis would allow QCA researchers to examine whether the solutions identified also hold up when other relevant control variables are entered along with these solutions into a regression model. Therefore, as an example, we included three control variables in our analyses ('model 2' in table 1). From these results it can be learned that, apparently, the added control variables do not appear to be extremely relevant for our main results. Solely in dataset 1, we find evidence that age affects implementation willingness. Solely in dataset 2, we find evidence that education slightly

affects implementation willingness. This shows that our results remain robust and implementation willingness appears to be less relevant on the added control variables than the powerfulness variables under study.

Table 1. OLS regression with implementation willingness as dependent variable

	Study 1		*Study 2*	
	Model 1	*Model 2*	*Model 1*	*Model 2*
SP	0.21**	0.21**	0.21**	0.21**
TP	0.17**	0.16**	0.04	0.04
OP	0.29**	0.29**	0.14**	0.15**
Age		-0.12*		-0.04
Tenure		0.03		0.01
Education		-0.02		-0.06*

**p<0.01; *p<0.05 Standardized coefficients (beta) are reported.

Second, we shed light on interpretation 2: Policy powerfulness interacts with policy meaningfulness. With regression analysis, we would model this as an interaction effect between powerfulness and meaningfulness. Table 2 shows the results, which shows that the interplay between policy powerfulness and meaningfulness does not strongly impact implementation willingness. Only the interaction between operational powerfulness and societal meaningfulness is statistically significant in dataset 1 and 2. Furthermore, the interaction between tactical powerfulness and client meaningfulness is significant in dataset 1, and the interaction between operational powerfulness and client meaningfulness in dataset 2. Please note that the standardized coefficients are relatively low; this indicates a – although statistically significant – weak relationship. Finally, the results indicate that societal and client meaningfulness correlate positively with implementation willingness. Clearly, more strongly than powerfulness and implementation willingness.

These findings partly align with our large large-N set-theoretic configurational analysis results: here, too, the empirical support for H3 is rather weak and has low explanatory power. However, an interaction effect has to be interpreted fundamentally differently from an INUS (Insufficient but necessary part of an unnecessary but sufficient) condition: the latter models a contingent effect, i.e., that the effect of one variable (here: meaningfulness) must unfold in combination with another variable (here: perceived discretion). The regression method applied is unsuitable to model this theoretical expectation.

Table 2. Simple regression results with implementation willingness as dependent variable and interaction of powerfulness with meaningfulness

	Study 1		*Study 2*	
	Model 1	*Model 2*	*Model 1*	*Model 2*
SP	0.10**	0.09**	0.07*	0.06
TP	0.12**	0.12**	0.04	0.04
OP	0.14**	0.13**	0.02	0.03
SM	0.26**	0.26**	0.18**	0.19**
CM	0.25**	0.27**	0.26**	0.26**
SP*SM		-0.03		0.07
TP*SM		0.07		0.01
OP*SM		-0.09*		-0.12**
SP*CM		0.02		-0.06
TP*CM		-0.08*		-0.01
OP*CM		0.03		0.10*

**p<0.01; *p<0.05. Standardized coefficients (beta) are reported. We used centralized variables for creating the interaction terms. As a check we also conducted the analyses with standardized instead of centralized variables and found the same results.

Alternative analysis 2: Polynomial regression

If the relationship between X and Y is not monotonic, a polynomial regression may do a much better job. In statistics, polynomial regression is a form of linear regression in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E(y |x), and has been used to describe nonlinear phenomena (although polynomial regression is technically a special case of multiple linear regression). There are three main arguments that indicate a linear relationship may not be a good model.

The first argument is theoretical: there are some relationships that a researcher will hypothesize is curvilinear. Clearly, if this is the case, including a polynomial term will be beneficial. However, necessity and sufficiency are not curvilinear relationships. With fuzzy sets, those cases support a set relation that cluster above or below the diagonal of the XY plot (Schneider and Wagemann 2012).

The second argument may result from visual inspection of your variables. A simple scatter plot can reveal a curvilinear relationship. However, the analyses of our data (scatterplots) do not conform such patters (see below).

The third argument may result from inspection of residuals. If you try to fit a linear model to curved data, a scatterplot of residuals (Y axis) on the predictor (X axis) will have patches of many positive residuals in the middle, but patches of negative residuals at either end (or vice versa). This is a good sign that a linear model is not appropriate, and a polynomial may do better. However, the analyses of our data (scatterplots of residuals) do not indicate such a pattern.

Polynomial results

We conducted a number of polynomial analyses with powerfulness and meaningfulness as the independent variables, and with implementation willingness as the dependent variable. The results emphasize that polynomial analysis does not provide us with a model that better fits the data than a simple linear regression model does (limited change in R Square). This is also shown by the scatterplots, which do not indicate a curvilinear relationship. Finally, the results of the polynomial analysis are not significantly better or easier to interpret than the linear results.