sociological science

Supplement to:

Miles, Andrew, Gordon Brett, Salwa Khan, and Yagana Samim. 2023. "Testing Models of Cognition and Action Using Response Conflict and Multinomial Processing Tree Models." Sociological Science 10: 118-149.

Appendix A: Studying Cognition with Implicit and Explicit Measures

Implicit measures can be used to tap into many types of automatic cognition sociologists care about (Miles 2019; Miles, Charron-Chénier, and Schleifer 2019), and so should allow scholars to assess automatic cognition as it relates to different types of behaviors using standard statistical techniques. For example, a researcher could measure the automatic and deliberate processes associated with a particular behavior using both implicit and explicit measures and then include them simultaneously in a regression model to see which explains the most variance. We suspect that this type of regression-based approach would be familiar and appealing to many sociologists who might therefore be hesitant to invest in learning response conflict tasks (RCTs) and multinomial processing tree (MPT) modeling. While it is true that a regression approach would provide some basic insights into the relative contributions of deliberate and automatic cognition to behavior, it also has significant limitations.

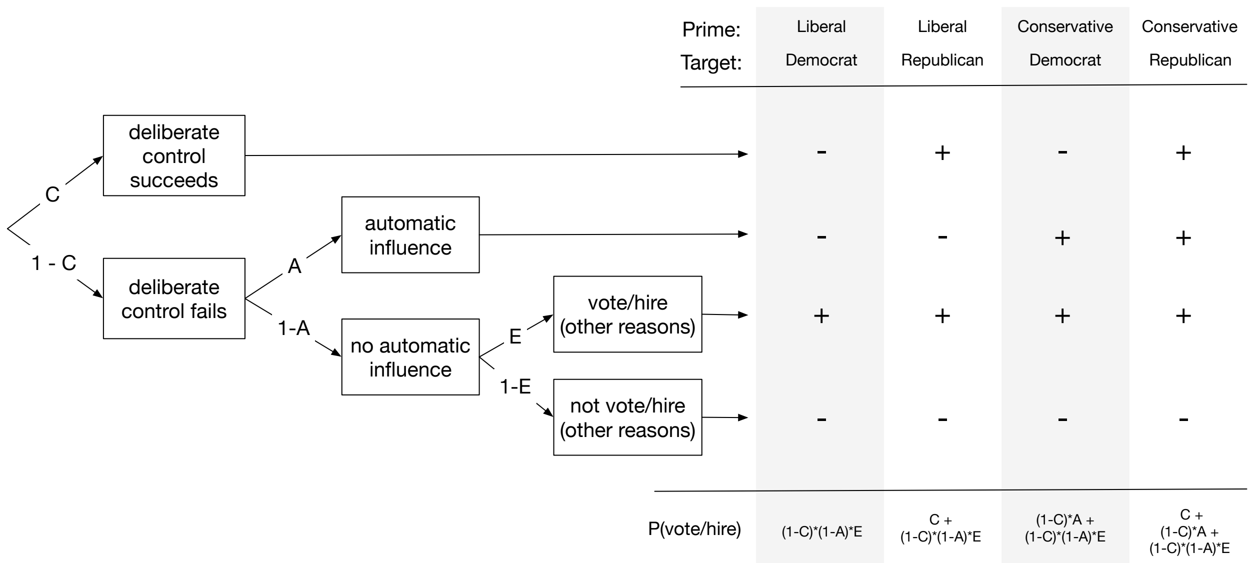
First, implicit and explicit measures capture *individual* constructs such as attitudes or mental schemas. While this makes them useful for determining if a particular type of construct is related to a behavior (e.g., do automatic associations of Black people with violence lead to avoidance behavior?), it also means that capturing the *total* influence of automatic or deliberate cognition would require having measures of *every* automatic or deliberate process relevant to that behavior. This is rarely feasible. Studying racially discriminatory behavior, for instance, might require measuring racial attitudes, personal identities, stereotypes, and habits both using explicit and implicit measures, an undertaking that would be both time-consuming and exhausting for respondents. In contrast, well-designed RCTs and MPT models capture the influence of all processes that produce an observed behavior.

Second, explicit and implicit measures are rarely process-pure, meaning that they could each tap into deliberate and automatic processing (Miles et al. 2019; Payne et al. 2010; Payne and Bishara 2009; Vila-Henninger 2015). On the plus side, including them both in a regression model should control away some of that shared variation and increase confidence that explicit measures capture deliberate cognition while implicit measures capture automatic cognition (Miles et al. 2019). However, the shared variation lost in this way might reflect influence by either automatic or deliberate cognition (or both). Regression analysis does not allow this variation to be attributed to one or the other type of processes, which could lead to an underestimation of how influential each type of process is. RCTs do not suffer from this problem, as they allow for estimates of cognitive processes that are process-pure (Yonelinas and Jacoby 2012).

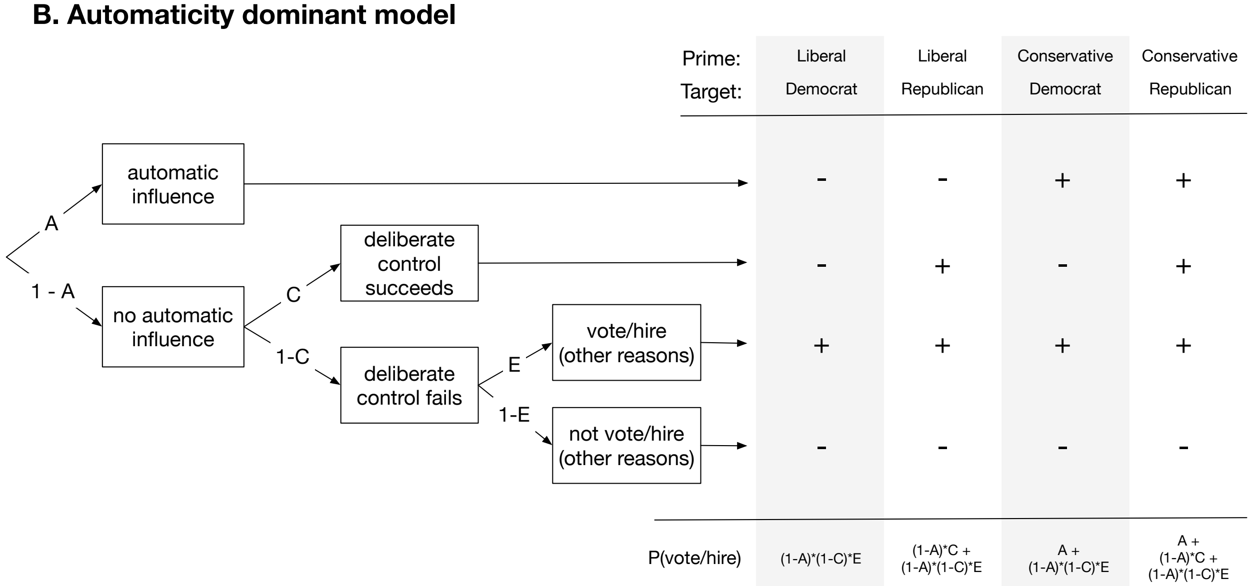
Finally, while comparing regression coefficients can give some sense for how important different cognitive processes are in shaping a behavior, regression analysis is not well suited to capturing sequences or dependencies of the sort implied by cognitive models (e.g., controlled processing is employed only when automatic processing fails to produce a response). This means that regression generally cannot be used to determine which cognitive processes are in control. MPT models, in contrast, build these type of control relationships directly into the model.

Appendix B: MPT models for conservatives in the politics sample from Miles et al. (2019)

A. Deliberation dominant model



B. Automaticity dominant model



Appendix C: MPT Models with Equivalent Fit

Sometimes two (or more) MPT models fit the data equally well, which makes it impossible to use fit statistics to adjudicate between them. Distinguishing between competing models is often an explicit research question, so it is important to understand what leads to equivalent fit and how to avoid it.

To determine when two models will fit the data equally well, we can:

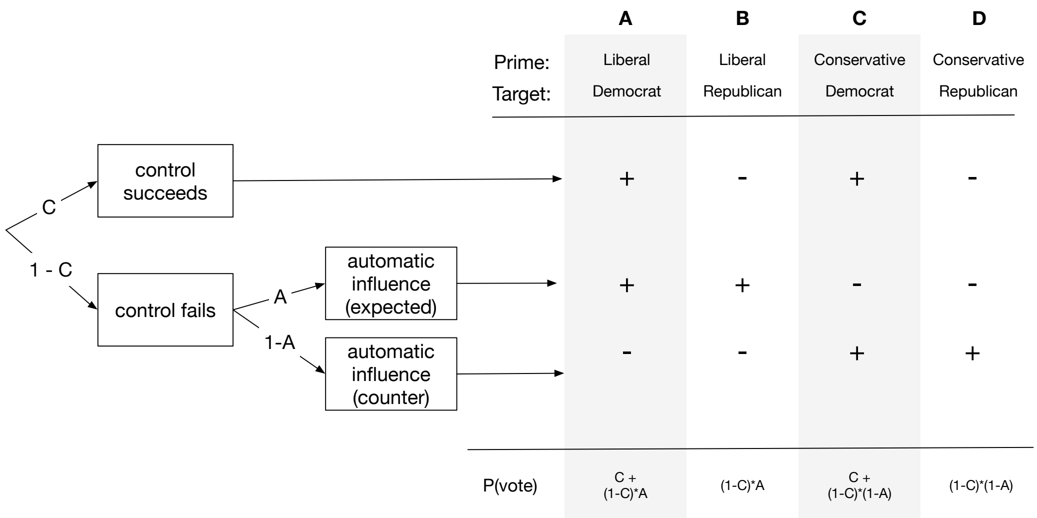
1. Write the parameters from the first model (call it model 1) in terms of the parameters from the other model (call it model 2)
2. Determine which values for parameters from model 2 return admissible values for all parameters from model 1. In MPT models, parameter estimates represent probabilities, and so are admissible if they are in the range of 0 to 1.

This approach determines whether estimated parameter values from one model can be represented by some configuration of the parameters from a different model.

This process is best illustrated with an example. We'll use the process dissociation and Stroop models described by Bishara and Payne (2009). Both models are quite simple, with only two parameters, A and C, which are assumed to represent the same quantities in both models (automatic and deliberate cognition). However, the two models differ in their structure. For convenience, the models are reproduced in Figure C1 below. We apply these models to the voting task data from the liberal subsample merely for convenience, as they will be familiar to those who have read the paper. Note that these models differ from models used in the main manuscript, which included three parameters—C, A, and E. This keeps the mathematics simpler while still illustrating the principles that likely explain why our analysis models have equivalent fit.

Figure C1: Process dissociation and Stroop models from Bishara and Payne (2009)

Process Dissociation Model



Stroop Model

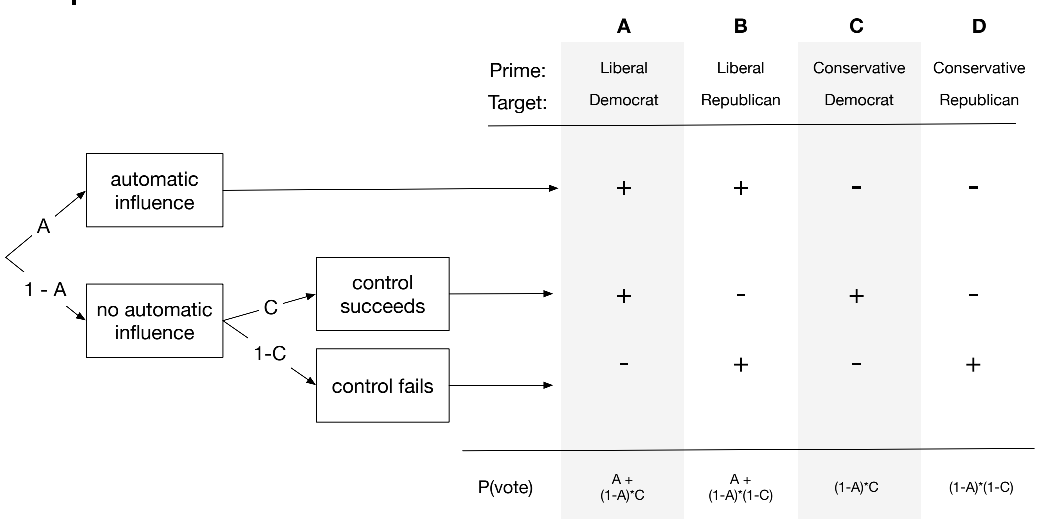


Figure C1 makes it straightforward to extract the model equations that define both the process dissociation and Stroop models. These are shown in Table C1 below. For simplicity, we replace the full notation for condition-specific probabilities such as p(vote | liberal, Democrat) with the condition-specific letter labels shown in Figure C1 (e.g., $p(A)$). To distinguish parameters from each model, we add a PD subscript to parameters from the process dissociation model, and an S subscript to parameters from the Stroop model. Each equation has also been given a number to make them easier to reference moving forward.

Table C1: MPT model equations for process dissociation and Stroop models

Eq.	Process Dissociation	Eq.	Stroop

1	$p(A) = C_{PD} + (1\text{-}C_{PD})A_{PD}$	**5**	$p(A) = A_S + (1\text{-}A_S)C_S$
2	$p(B) = (1\text{-}C_{PD})A_{PD}$	**6**	$p(B) = A_S + (1\text{-}A_S)(1\text{-}C_S)$
3	$p(C) = C_{PD} + (1\text{-}C_{PD})(1\text{-}A_{PD})$	**7**	$p(C) = (1\text{-}A_S)\, C_S$
4	$p(D) = (1\text{-}C_{PD})(1\text{-}A_{PD})$	**8**	$p(D) = (1\text{-}A_S)(1\text{-}C_S)$

Notice that both equation 1 from the process dissociation model and equation 5 from the Stroop model include *p*(*A*). Likewise, equations 2 and 6 both contain *p*(*B*), equations 3 and 7 contain *p*(*C*), and equations 4 and 8 contain *p*(*D*). These shared terms make it possible to express the parameters from one model in terms of parameters from the other. We will first focus on writing C_{PD} and A_{PD} from the process dissociation model in terms of A_S and C_S from the Stroop model.

Let's begin by rearranging equation 1. Solving for A_{PD} gives us:

$$A_{PD} = \frac{p(A) - C_{PD}}{1 - C_{PD}}$$

Substituting this into equation 2 and simplifying allows us to write *p(B)* in terms of *p*(*A*) and C_{PD}:

$$p(B) = (1 - C_{PD})A_{PD}$$
$$p(B) = (1 - C_{PD})\left(\frac{p(A) - C_{PD}}{1 - C_{PD}}\right)$$
$$p(B) = p(A) - C_{PD}$$

Equation 6 from the Stroop model also includes *p*(*B*). We can make a substitution, and rearrange to solve for C_{PD}.

$$p(B) = A_S + (1 - A_S)(1 - C_S)$$
$$p(A) - C_{PD} = A_S + (1 - A_S)(1 - C_S)$$
$$C_{PD} = -A_S - (1 - A_S)(1 - C_S) + p(A)$$

We now have an equation that expresses C_{PD} in terms of A_S, C_S, and *p*(*A*). Equation 5 gives us a substitution for *p*(*A*). Then we simplify.

$$C_{PD} = -A_S - (1 - A_S)(1 - C_S) + A_S + (1 - A_S)C_S$$
$$C_{PD} = -(1 - A_S)(1 - C_S) + (1 - A_S)C_S$$
$$C_{PD} = (1 - A_S)(2C_S - 1)$$

(E1)

Equation E1 now expresses C_{PD} entirely in terms of parameters from the Stroop model.

We can use the same substitution approach for A_{PD}. We will begin by substituting equation 6 in for $p(B)$ in equation 2.

$$A_S + (1 - A_S)(1 - C_S) = (1 - C_{PD})A_{PD}$$

We can now use equation E1 to substitute in for C_{PD}, then solve for A_{PD} and simplify.

$$A_S + (1 - A_S)(1 - C_S) = \big(1 - (1 - A_S)(2C_S - 1)\big)A_{PD}$$

$$A_{PD} = \frac{A_S + (1 - A_S)(1 - C_S)}{1 - (1 - A_S)(2C_S - 1)}$$

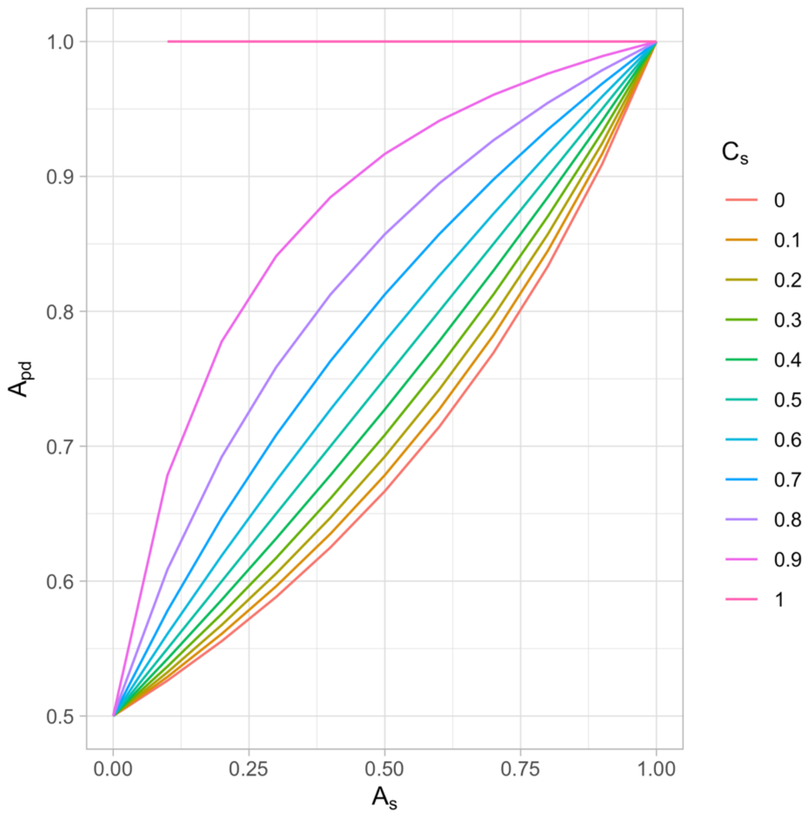
$$A_{PD} = \frac{1 + C_S(A_S - 1)}{1 - (1 - A_S)(2C_S - 1)} \tag{E2}$$

At this point, we have equations E1 and E2 that express C_{PD} and A_{PD} entirely in terms of the parameters A_S and C_S from the Stroop model.

We can now use these equations to determine where the process dissociation and Stroop models will have equivalent fit to the data. Equivalent fit will occur whenever values for A_S and C_S produce admissible values for C_{PD} and A_{PD} using equations E1 and E2. Recall that admissible values must lie between 0 and 1.

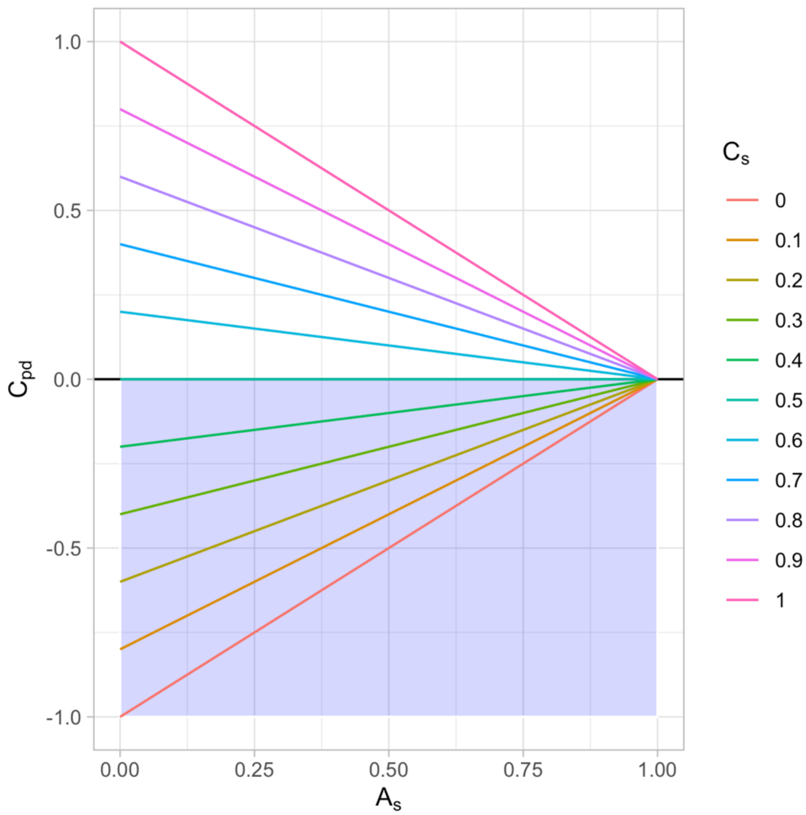
Figure C2 shows how A_{PD} varies at values of A_S and C_S ranging from 0 to 1 (with increments of 0.10). Estimates of A_{PD} are bounded below by 0.50 and above by 1. All possible values of A_{PD} are admissible, so equation E2 places no constraints on the values at which the process dissociation and Stroop models will have equivalent fit (the exception being when A_S is 0 and C_S is 1, in which case is A_{PD} undefined).

Figure C2: Predicted variation in A_{PD} over the parameter space of A_S and C_S



Of course, equal fit also requires that values of C_{PD} are admissible. Figure C3 shows values of C_{PD} when A_S and C_S range from 0 to 1 (with increments of 0.10). In this case, admissible values for C_{PD} are only produced when $C_S \geq 0.50$. Values of C_S below 0.50 produce negative estimates for C_{PD} (shown in blue).

Figure C3: Predicted variation in C_{PD} over the parameter space of A_S and C_S



In sum, Figures C2 and C3 indicate that the process dissociation model can achieve equivalent fit to the Stroop model when the C_S parameter is at or above 0.50, regardless of the value of the A_S parameter. Likewise, it cannot achieve equivalent fit when the C_S parameter is below 0.50, regardless of the value of the A_S parameter. Practically, these results imply that if a behavior is produced by the Stroop model, a comparison of the Stroop model with the process dissociation model will only be able to identify the Stroop model as correct if the data produce a C_S parameter below 0.50.

We turn now to the case where the process dissociation model is correct. To see what the parameter space looks like for identifying a correct process dissociation model, we can follow the same process as above, but this time express A_S and C_S in terms of A_{PD} and C_{PD}. For brevity we will skip straight to the final equations:

$$A_S = (1 - C_{PD})(2A_{PD} - 1) \tag{E3}$$

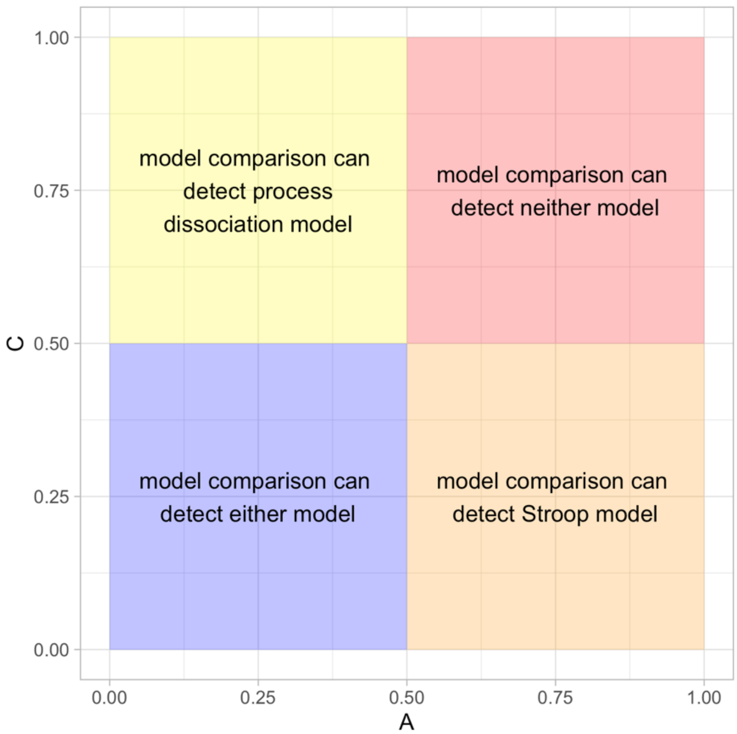
$$C_S = \frac{1 + A_{PD}(C_{PD} - 1)}{1 - (1 - C_{PD})(2A_{PD} - 1)} \tag{E4}$$

The process dissociation model and Stroop model will have identical fit to the data whenever values for C_{PD} and A_{PD} produce admissible values for A_S and C_S using equations E3 and E4. As before, determining where values for C_{PD} and A_{PD} produce *inadmissible* values for A_S and C_S will delineate the parameter space where it will be possible to determine when a process dissociation model is correct.

It turns out that equation E4 places no limitations on the parameter space, but equation E3 only produces admissible values for A_S if $A_{PD} \geq 0.50$.

Collectively, equations E1 to E4 thus allow us to define the parameter space in which it is possible to detect whether either a process dissociation model or a Stroop model is correct. This space is shown in Figure C4. The yellow and blue areas represent the values of *A* and *C* at which a process dissociation MPT model can accurately detect data that are generated by an underlying process dissociation model. The orange and blue areas show where a Stroop MPT model can accurately detect data generated by a Stroop type of process. The red area represents the values of *A* and *C* at which both MPT models will have equivalent fit regardless of what the true underlying model is, and hence be uninformative. Only parameters in the blue area can accurately detect data generated by *either* a process dissociation or a Stroop type of process. Practically, this suggests that researchers interested in comparing the process dissociation and Stroop models should design studies that produce probabilities of relying on deliberate (*C*) or automatic (*A*) cognition that fall below 0.50.

Figure C4: Parameter space where either a process dissociation or Stroop model can be detected as correct



Demonstration Using Simulation

To illustrate these principles, we can generate data that is consistent with either a process dissociation or Stroop model with values of *C* and *A* drawn from the four quadrants shown in Figure C4. We can then fit both models to those data to determine if our predictions are accurate.

We generated data according to the process dissociation and Stroop models using the equations in Table C1. For each model, we set parameters to every unique combination of C = {0.3, 0.7} and A = {0.3, 0.7}, for four combinations in total, one each for each quadrant shown in Figure C4. We then used the resulting probabilities to generate n = 300 in each condition (shown as conditions A, B, C, and D in Figure C1). To avoid perfect fits, we then added a randomly generated count ranging between -15 and 15 (i.e, -n/20 to n/20). This resulted in eight data sets in total: four based on the process dissociation model, and four based on the Stroop model.

Next, we fit both the process dissociation and Stroop model to each dataset and extracted the goodness of fit statistic G^2 from each model. Results are shown in tables C2 and C3. Table C2 presents results from datasets generated using the Stroop model. Consistent with Figure C4, the Stroop model was correctly detected when the data were generated using C = 0.3, as evidenced by the much smaller G^2 statistics compared to the process dissociation model (shown in bold). However, G^2 statistics were identical across the process dissociation and Stroop models when using C = 0.7.

Table C2. Fit statistics of models fit to data simulated from Stroop model

		Process Dissociation		Stroop	
C	A	G^2	p-value	G^2	p-value
0.3	0.3	87.62	<0.001	**0.12**	**0.940**
0.3	0.7	38.73	<0.001	**7.27**	**0.026**
0.7	0.3	0.08	0.962	0.08	0.962
0.7	0.7	9.89	0.007	9.89	0.007

Note: the smallest G^2 statistic is bolded for each simulated data set

Table C3 presents results from models fit to data generated using the process dissociation model. We see that the process dissociation model had superior fit to the data compared to the Stroop model only when A = 0.3. When A = 0.7, fit between the two models was identical.

Table C3. Fit statistics of models fit to data simulated from process dissociation model

		Process Dissociation		Stroop	
C	A	G^2	p-value	G^2	p-value
0.3	0.3	**0.32**	**0.854**	128.28	<0.001
0.3	0.7	4.92	0.085	4.92	0.085
0.7	0.3	**4.54**	**0.103**	52.78	<0.001
0.7	0.7	4.56	0.102	4.56	0.102

Note: the smallest G^2 statistic is bolded for each simulated data set

Looking across Tables C2 and C3, we see that only when C = 0.3 and A = 0.3 could we identify the correct model both when it was produced by an underlying process dissociation and when it was produced by an underlying Stroop model, just as indicated by Figure C4.

Moving Beyond Simple Models

The process dissociation and Stroop models are both simple models, which makes them ideal for demonstration purposes. In practice, researchers will likely want to fit more complex models. The approach demonstrated here—while more tedious with more parameters—can be used to identify the parameter space in which models will have equivalent fit, and hence where model comparisons will be uninformative.

In general, equivalent fit should be harder to achieve as the number of parameters increases. Achieving equal fit for two models that each have four parameters, for instance, would require that parameter values from one model have corresponding admissible values for all four of the parameters in the other model. Thus, adjudicating between competing models should be easier as the number of parameters in each model increases.

Equivalent Fit for the Models in the Main Manuscript

The foregoing discussion indicates that equivalent fit is determined both by characteristics of the data (i.e., the true underlying values of the parameters) and by the complexity of the models fit. While we do not know the true values of the parameters that describe the processes involved in the voting and hiring tasks analyzed in the main manuscript, we do know that the models were relatively simple, with only one parameter more than the models described in this appendix. A reasonable conclusion is therefore that these models were too simple to capture whatever differences exist in the data that would have allowed us to determine whether deliberate or automatic processes control behavior in cases where the two conflict. Alternately, the data themselves might not be informative about issues of control (comparable to the red quadrant shown in Figure C4).

Appendix D: Evaluating Whether Primes Have Variable Effects

The primes used by Miles et al. (2019) consisted of liberal/conservative prime words, images of harmful or caring behavior, and pictures of black and white faces. The models in the paper estimated a single A parameter, which implicitly assumes that the primes used in a given sample had the same probability of guiding respondents' behavior. This might not be the case. For instance, some primes might have been more likely to be experienced as negative, and negative stimuli could have stronger effects than positive stimuli (c.f., Baumeister et al. 2001). If primes have different effects, estimating a single A parameter will produce a biased estimate of the influence of prime-related automatic processes.

We can account for this by estimating separate A parameters for different kinds of primes, e.g., A_1 for liberal primes, and A_2 for conservative primes. However, adding an additional A parameter saturates all models. This makes it impossible to calculate fit statistics, and in some cases can lead to estimation difficulties. Nonetheless, observing the estimates from models that include both A_1 and A_2 might provide insight into whether priming effects differ across prime types, and hence whether the results presented in the main manuscript are likely to be biased because they include only a single A parameter.

Estimates from deliberation-dominant and automaticity-dominant models with A_1 and A_2 are shown in Table D1. Deliberation-dominant models in the two politics and morality samples encountered estimation difficulties, and their estimates might not be reliable. However, the deliberation-dominant model in the race sample as well as all automaticity-dominant models estimated without difficulty. If we consider just these models, there is a clear pattern: estimates of A_1 and A_2 either do not differ (in the race sample), or do not differ by much. This suggests that estimating models with a single A parameter is unlikely to bias estimates of prime-related automatic influence to any substantial degree.

Table D1: Parameter estimates from models with separate A parameters for different prime types

	Deliberation-Dominant				Automaticity-Dominant		
	Est.	**95% CI (Lower)**	**95% CI (Upper)**		**Est.**	**95% CI (Lower)**	**95% CI (Upper)**
1. Politics (Liberal)							
A_1	0.24	---	---	**A_1**	0.10	0.05	0.15
A_2	0.00	---	---	**A_2**	0.17	0.10	0.24
C	0.29	---	---	**C**	0.33	0.31	0.35
E	0.25	---	---	**E**	0.33	0.29	0.38
2. Politics (Conservative)							
A_1	0.00	-49234.58	49234.58	**A_1**	0.09	-0.02	0.20
A_2	0.25	-15015.68	15016.17	**A_2**	0.14	0.07	0.20
C	0.33	0.30	0.36	**C**	0.37	0.34	0.41
E	0.29	-14204.17	14204.75	**E**	0.33	0.26	0.40
3. Morality							
A_1	0.00	-1525.46	1525.46	**A_1**	0.05	0.02	0.09

A_2	0.09	-866.27	866.46	**A_2**	0.01	-0.02	0.04
C	0.50	0.49	0.52	**C**	0.52	0.50	0.53
E	0.62	-586.63	587.86	**E**	0.57	0.54	0.60
4. Race							
A_1	0.00	-0.24	0.24	**A_1**	0.00	-0.04	0.04
A_2	0.00	-0.17	0.17	**A_2**	0.00	-0.03	0.03
C	0.47	0.45	0.48	**C**	0.47	0.45	0.48
E	0.59	0.49	0.68	**E**	0.59	0.56	0.62

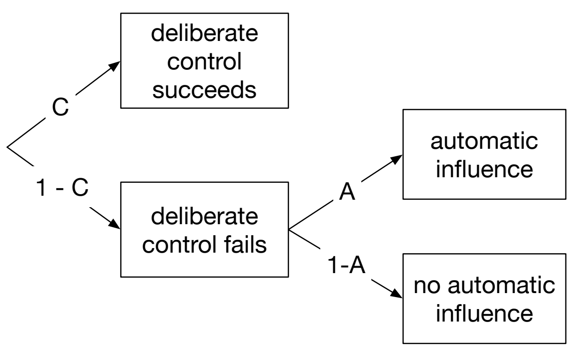
Appendix E: Designing a Response Conflict Study

In this appendix, we offer our step-by-step recommendations for how to design an informative response conflict study of cognition.

1. Sketch the basic MPT model you want to test

Many cognitive theories are expressed verbally, but this can be imprecise. Drawing a tree diagram of the cognitive model will help to clarify issues such as a) how many processes there are, and b) how those processes relate to one another to produce a behavior. For example, the process dissociation model shown in Figure E1 has two processes, deliberate cognition (C) and automatic cognition (A) and specifies that automatic processes only control behavior if deliberate processes fail to do so.

Figure E1: The process dissociation model

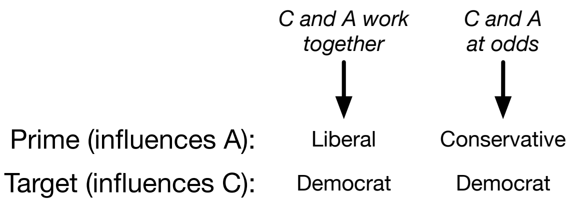


2. Determine how you can manipulate the processes in your MPT model

An MPT model's ability to isolate cognitive processes depends on having observable variation in those processes. The raw data for the model is how often people perform a behavior under conditions that vary the extent to which they rely on each process. Hence, you must create conditions in which the cognitive processes specified in your model work alternately in tandem and at odds—that is, you must create response conflict data (or find where it naturally occurs).

For example, the voting task used in Miles et al. (2019) uses liberal and conservative primes to encourage automatic responses that favor or oppose voting for candidates, as shown in Figure E2. When the candidate matches a respondent's own political party, a favorable automatic influence works in tandem with the deliberate intention to vote for the candidate. However, an "oppose" automatic influence works against that intention.

Figure E2: Illustration of processes working in tandem and at odds to produce voting behavior

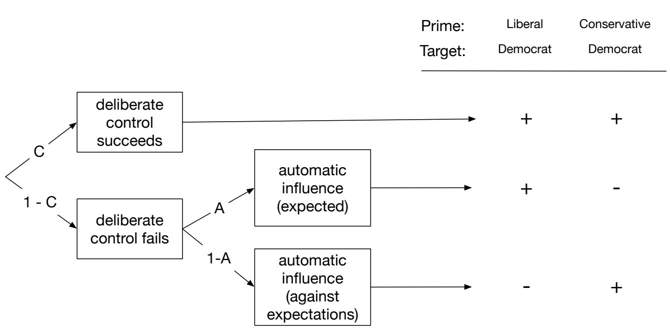


You can manipulate processes either within-persons or between-persons. For example, the voting task from Miles et al. (2019) used a within-person approach in which respondents received both party-consistent and party-inconsistent primes. A between-person variation would be to randomly assign participants to experience *either* party-consistent or party-inconsistent primes, but not both. Generally, within-person approaches are preferable because they capture variation in how processes operate for each individual, which makes it possible to fit individual MPT models and hence calculate individual-level estimates. However, a between-person approach might be preferable under certain circumstances, such as when researchers want to capture single instances of real-world behavior (e.g., performing a prosocial act).

3. Use the MPT model to make predictions

Use the MPT model to predict how respondents will behave in each condition. For example, let's assume we are going to fit a process dissociation model to data from a sample of liberals. Our predictions are shown in Figure E3. In the liberal prime/Democrat candidate condition, we would expect the deliberate intention and the prime-related automatic reaction to both produce a "vote for" response (shown as a + in the diagram below). According to our model, the only way to produce a "vote against" response in this condition would be if the automatic influence ran counter to expectations. In the conservative prime/Democrat candidate condition, however, voting for a Democrat would only occur if the respondent was able to carry through on her intentions, or else she had an unexpectedly positive reaction to the conservative prime.

Figure E3: Predicted voting behavior of liberal respondents during a voting task



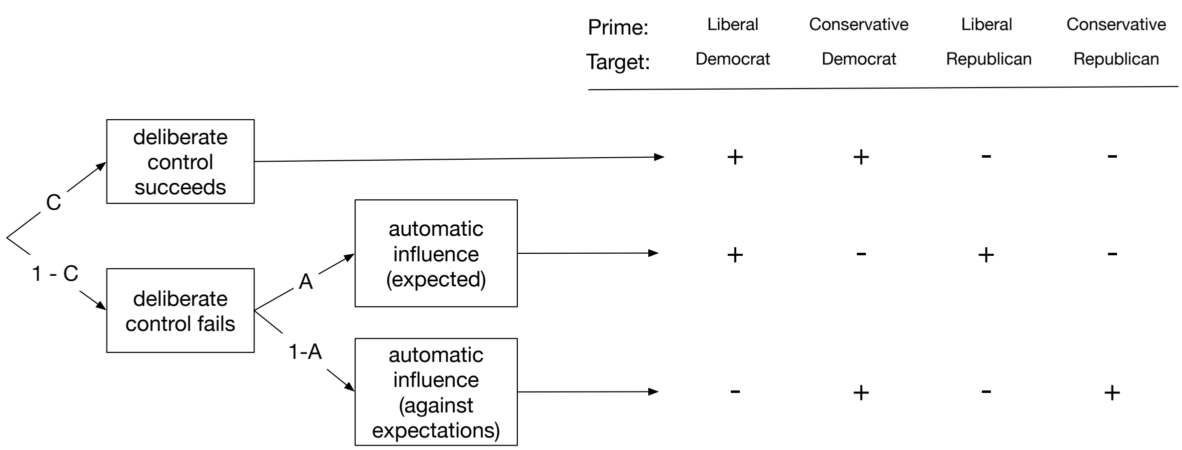
Note: + indicates a vote for the target candidate.

4. Modify the RCT and/or MPT if needed

As you develop an RCT that allows you to test your MPT, you might find that your first draft does not give you everything you need. For instance, our example in Figure E3 shows an RCT with two conditions. That means that our data provide two degrees of freedom (*d.f.*) for analysis. Our model has two parameters and therefore fully uses up the available information. Because our model is saturated, we will be able to obtain parameter estimates but we will not be able to compute the G^2 fit statistic. We also cannot add any parameters to our model to capture other processes that might be influencing respondents' decisions.

The solution is to add more conditions. In this case, we can add Republican candidates as additional target stimuli to the model shown in Figure E3, which produces the model shown in Figure E4. This now gives us four unique prime/candidate combinations, and hence four conditions and four *d.f.* Of course, anytime we add conditions we must also ask whether we need additional parameters to fully capture the processes involved. For example, do we believe that liberal respondents will react equally strongly to both liberal and conservative primes? If so, we can estimate those effects with a single A parameter. If not, we should estimate a separate A parameter for each type of prime (see Appendix D).

Figure E4: An expanded MPT model for predicted voting behavior of liberal respondents during a voting task



Note: + indicates that a vote for the target candidate.

In general, the key to adding more conditions is to think about how you can modify the RCT in a way that will affect just one of your parameters. For example, you might further modify the design shown in Figure E4 by varying how much time respondents have to vote. Because imposing a shorter response deadline should reduce the probability of relying on deliberate processes but should not affect automatic processes, you would need to include a separate C parameter in each response time condition, but no additional A parameters. Thus, the full model would have four parameters: two C parameters (one for each response time condition) and two A parameters (one for each type of prime). However, the total number of conditions would be the unique combinations of prime X candidate X response time. That would be 2 prime types X 2 candidate types X 2 response times = 8 conditions. The data would thus provide 8 *d.f.* for analysis.

If you use a within-person design for your RCT, you may also want to assess whether you will have sufficient sample sizes for individual-level analyses. Because the raw data for MPT models are counts of how often respondents perform the behavior in each unique condition, you must ensure that a) respondents have many chances to perform the behavior under each unique set of conditions, and b) the behavior occurs at least some of the time in every study condition—this will ensure that there are no empty cells of data which could lead to estimation problems. For example, individual level analyses of the model shown in Figure E4 would require that each respondent has the chance to vote for both Democrats and Republicans many times following both liberal and conservative primes. Further, the task should be designed so that respondents cannot perfectly follow through on their intentions to vote only for party-consistent candidates. This might be accomplished by using unfamiliar candidates (so that respondents sometimes forget which party they represent), or by asking respondents to answer rapidly.

If the purpose of your research is to compare the validity of different MPT models—e.g., to determine whether an automaticity- or deliberation-dominant model better represents the data—then you must guard against the possibility that your models will have equivalent fit to the data and hence be unable to answer your research question. See Appendix C for an explanation of how to predict when this will occur so that modifications to your study design can be made prior to data collection.

5. Validate MPT model parameters

Nothing in an MPT model guarantees that the parameters mean what you think they mean. MPT parameters capture latent processes, and hence they are not directly observable. This means that the interpretation of these parameters must be validated in some other way.

Typically, validation work is accomplished by altering the conditions under which an RCT is performed and observing whether model estimates respond in theoretically expected ways. For example, dual-process theory suggests that deliberate processes rely heavily on cognitive resources such as attention, while automatic processes do not. Hence, we can test whether C and A represent what we think they represent by having respondents complete our RCT while distracting or loading their conscious minds. If the C parameter really captures deliberate cognition, it should diminish when cognitive resources are depleted, but the A parameter should remain unchanged.

Addendum: Using naturally occurring data

RCT data most often comes from custom-designed tasks, but this need not be the case. Naturally occurring data could also be used so long as conditions can be found where deliberate and automatic processes operate in tandem or at odds with one another.

For example, the relative strength of automatic and deliberate processes in producing behavior consistent with traditional gender norms might be assessed by observing individuals in professional and private settings. Automatic influences should not change across contexts, but intentions about how to act might. Similarly, the role of racial bias in hiring might be assessed by fitting MPT models to counts of how often managers hire equally qualified Black and white candidates. The assumption here would be that managers intend to hire qualified candidates, but that racial biases activate automatic processes in ways that work for or against this intention.

References

Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs. 2001. “Bad Is Stronger Than Good.” *Review of General Psychology* 5(4):323–70.

Bishara, Anthony J. and B. Keith Payne. 2009. “Multinomial Process Tree Models of Control and Automaticity in Weapon Misidentification.” *Journal of Experimental Social Psychology* 45(3):524–34.

Miles, Andrew. 2019. “An Assessment of Methods for Measuring Automatic Cognition.” Pp. 341–66 in *The Oxford Handbook of Cognitive Sociology*, edited by W. Brekhus and G. Ignatow. New York: Oxford University Press.

Miles, Andrew, Raphaël Charron-Chénier, and Cyrus Schleifer. 2019. “Measuring Automatic Cognition: Advancing Dual-Process Research in Sociology.” *American Sociological Review* 84(2):308–33.

Payne, B. Keith and Anthony J. Bishara. 2009. “An Integrative Review of Process Dissociation and Related Models in Social Cognition.” *European Review of Social Psychology* 20(1):272–314.

Payne, B. Keith, Deborah L. Hall, C. Daryl Cameron, and Anthony J. Bishara. 2010. “A Process Model of Affect Misattribution.” *Personality and Social Psychology Bulletin* 36(10):1397–1408.

Vila-Henninger, Luis Antonio. 2015. “Toward Defining the Causal Role of Consciousness: Using Models of Memory and Moral Judgment from Cognitive Neuroscience to Expand the Sociological Dual-Process Model.” *Journal for the Theory of Social Behaviour* 45(2):238–60.

Yonelinas, Andrew and Larry Jacoby. 2012. “The Process-Dissociation Approach Two Decades Later: Convergence, Boundary Conditions, and New Directions.” *Memory & Cognition* 40(5):663–80.