How to Map Theory: Reliable Methods Are Fruitless Without Rigorous Theory

Kurt Gray
University of North Carolina, Chapel Hill

Abstract
Good science requires both reliable methods and rigorous theory. Theory allows us to build a unified structure of knowledge, to connect the dots of individual studies and reveal the bigger picture. Some have criticized the proliferation of pet “Theories,” but generic “theory” is essential to healthy science, because questions of theory are ultimately those of validity. Although reliable methods and rigorous theory are synergistic, Action Identification suggests psychological tension between them: the more we focus on methodological details, the less we notice the broader connections. Therefore, psychology needs to supplement training in methods (how to design studies and analyze data) with training in theory (how to connect studies and synthesize ideas). This article provides a technique for visually outlining theory: theory mapping. Theory mapping contains five elements, which are illustrated with moral judgment and with cars. Also included are 15 additional theory maps provided by experts in emotion, culture, priming, power, stress, ideology, morality, marketing, decision-making, and more (see all at theorymaps.org). Theory mapping provides both precision and synthesis, which helps to resolve arguments, prevent redundancies, assess the theoretical contribution of papers, and evaluate the likelihood of surprising effects.

Keywords
action identification, scientific methodology, social cognition, morality, theory

The trick to forgetting the big picture is to look at everything close up.

—Chuck Palahniuk

What makes for good science? Test your intuitions by deciding which of these studies makes a more important contribution:

Study 1 involves 123 White, male, middle-class university students who respond to a few multiple-choice questions. Researchers report only descriptive statistics, do not preregister their hypotheses, and do not share their data.

Study 2 recruits a representative sample of over 30,000 people across multiple countries who complete a broad sample of explicit and implicit measures. Researchers report descriptive and inferential statistics and use both factor analyses and structural equation modeling. They preregister their hypotheses and make their data publicly available.

Most psychologists would rate Study 2 as more important than Study 1 because its methods are more reliable, with higher power, more diverse participants, more varied items, more comprehensive analyses, and more transparency. But I would argue that Study 1 is rather important: It was conducted by Solomon Asch (1951) and revealed that people will give the wrong answer to a very obvious question—Which line is longest?—in order to conform. Despite having methodological features that many modern scientists would condemn, this study made an important and enduring contribution to psychological science. What of Study 2? Its findings might be important, or trivial, or perhaps even misleading. The point is that you cannot judge science by its methods alone.

Although good science must meet a threshold of reliability, its value hinges on its relevance to pressing social problems and to theory—the broader structure of

Corresponding Author:
Kurt Gray, Department of Psychology and Neuroscience, CB 3270, University of North Carolina, Chapel Hill, NC 27510.
E-mail: kurtgray@gmail.com

This article was supposed to have been published in the July 2017 Special Symposium on the Future Direction of Psychological Science. It inadvertently was not included, for which the journal apologizes.

—Bob Sternberg, Editor
Of Bricks and Buildings

Doing science involves building a structure of knowledge, but with uneven bricks. Each brick is a single study, with its size and shape set by the knowledge it reveals. As masons know, strong structures need strong bricks, and the same is true in science. Just as a few weak bricks can bring down an entire wall, unreliable studies can threaten entire subfields. The importance of reliable methods in science is therefore indisputable and recommends increased statistical power, improved transparency, and appropriate analyses.

However, strong bricks are not enough for good buildings; instead, they must be combined precisely. As Henri Poincaré (1905) wrote, “science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house” (p. 141). Exhortations to increase reliability sometimes forget that individual studies matter only to the extent that they contribute to a unified structure of understanding, whether by inspiring new work, creating new connections, prompting reinterpretations of past work, and facilitating new interventions.

The quest for reliable research methods—for making good bricks—is certainly noble, but the mere collection of reliable studies does not make for good science. We must remember that we scientists are not only brick-makers but also architects and turn our attention back to building—to theory.

The Benefits of Theory

The role of theory in science is controversial, with some believing that there is often too much of it. Mischel (2008) suggests that theories in psychology are like toothbrushes: Everyone wants their own, and no one wants to use others’. It’s true that psychologists delight in labeling effects and ideas, but so do all scientists: Physicists name particles, chemists name reactions, and biologists name species. Crucially, naming phenomena as we name our children—so that they become ours—is not “theory” but are “Theories,” capitalized like all proper nouns, and pluralized to reflect their multitude. Theories need no championing, but lowercase, generic “theory” does. Theory specifies the interconnection of knowledge; it is the blueprints of the building, the guide for making new bricks, and the mortar that holds them together.

A more classic metaphor for theory is a web, or nomological net (Cronbach & Meehl, 1955), that outlines constructs and their operationalizations and then explores their interconnections. Central to nomological nets are questions of validity: content validity (Does a measure actually assess the construct?), convergent validity (Is a construct associated with similar constructs?), and divergent validity (Is a construct distinct from dissimilar constructs?). Questions of theory are therefore ultimately questions of validity, as both focus on the interconnection between findings. The only difference between them is in popular usage, with “validity” referring to only a single claim and “theory” referring to a set of claims—to the broader picture.

Understanding theory as large-scale validity underscores its importance. By connecting scientific findings, it better defines them, reveals gaps in knowledge, and links disparate disciplines findings—just consider the integrative power of the theory of evolution. One example in psychology is the theories of social cognition that link social judgment to cognitive categorization (Fiske & Taylor, 2013). These theories give psychology new understandings of stereotyping and prejudice, allowing for new interventions and countless new research avenues.

The parallels between theory and validity provide the insight that reliability is necessary but not sufficient for theory. As hundreds of undergraduate textbooks assert, validity requires a threshold of reliability. Without a minimum association with itself (e.g., internal consistency), a measure cannot be associated with others. But even reliable measures can lack validity and fail to accurately assesses what you believe they should. Simply making a brick does not guarantee that it fits well into the structure of science.

The Tension Between Methods and Theory

Reliable methods (i.e., reliability) and rigorous theories (i.e., validity) are synergistic, as new methods can reveal new knowledge (Greenwald, 2012) that shape theory. For example, theories of social cognition were transformed by measures of implicit attitudes (Greenwald, Nosek, & Banaji, 2003; Payne, Cheng, Govorun, & Stewart, 2005). In his survey of the Nobel awards, Greenwald found two kinds of “method–theory synergy”: (a) Theories suggest new methods, and (b) new methods generate previously inconceivable data, inspiring previously inconceivable theories. Despite this overarching synergy, one idea suggests a tension between them within the minds of individual researchers: the theory of action identification.

Action identification suggests that actions are inherently ambiguous and lead to multiple “identifications” (Vallacher & Wegner, 1987). For example, shooting a gun could be merely moving your finger, or starting a revolution. Identifications are called low level if they focus on the close-up concrete details of the action—on the “how”—and are
called high level if they focus on the big-picture abstract reasons for the action—on the “why.” Research reveals that these two levels of identification often conflict (Vallacher & Wegner, 1987), with a focus on the “how” reducing an appreciation of the “why.” For example, identifying getting married as “booking a photographer” makes it less about “expressing your love.”

Applying action identification to science suggests that the more we focus upon sample sizes and data transparency (the bow of science), the less we focus upon theory building (the why of science). A similar prediction is made from the construal level theory (Trope & Liberman, 2011), which suggests that the more you focus on the details within individual studies, the more you lose sight of the connections between these studies, missing the forest for the trees. In the language of Cronbach and Meehl, the more science focuses upon reliability—Do these effects persist?—the less it focuses upon validity—What exactly do these effects mean?

The tension between “how” versus “why” may be hard to escape, but it needs to be managed, as both reliable methods and rigorous theory are essential to good science. We must embrace pluralism and respect scientists who focus primarily upon either developing reliable methods or building rigorous theory. Such scientific specialization is natural, as when physicists concentrate on either theory or experiments.

Scientific pluralism also involves teaching graduate students about both methods and theory. There are already many courses about reliable methods (most programs have a “methods” course), but there unfortunately is little explicit training in theory. This lack of theory-based instruction may stem from action identification, as it seems strange to specify the “how” of understanding “why.” Nevertheless, there is a concrete way to explore abstract ideas: theory mapping.

**Theory Mapping**

Theory mapping involves drawing out links between constructs, albeit in a specific way. Given that it focuses upon larger scale associations rather than causal connections, theory maps are meant to complement the diagrams of mediation models and Structural Equation Modeling, not replace them. It is important to note that this technique is preliminary and it should be modified by researchers to fit their own purposes. It is also worth acknowledging that—as with any visual figure—comprehensiveness must be balanced with readability, and so researchers must choose which theoretical links are most important to map. Despite these caveats, theory mapping allows the concrete display of knowledge structures, which helps with theory evaluation (Gawronski & Bodenhausen, 2015) and encourages the features of “high-quality science” (Finkel, Eastwick, & Reis, in press). It also allows newcomers to a field to see the most relevant constructs at a glance.

I will illustrate theory mapping in two ways: with cars and with moral psychology. Cars have nothing to do with psychology but intuitively illustrate theory mapping’s principles. A more technical application of the technique is provided by my work on dyadic morality, which suggests that we make moral judgments by comparing acts to a cognitive template of perceived harm (for a review, see Schein & Gray, 2017). This harm is “dyadic,” involving an intentional agent (a perpetrator) causing damage to a vulnerable patient (a victim), often represented as “A→P.”

Dyadic morality explains why acts like murder and assault are more immoral than tax evasion and double parking (they involve more perceived harm; Schein & Gray, 2015). It also explains why cultures moralize values like purity and loyalty to the extent that they view their violation as harmful (Shweder, Much, Mahapatra, & Park, 1997). For example, if you think gay marriage destroys both souls and society—like many conservative evangelicals—then you see it as immoral. Conversely, if you think gay marriage harms no one—like many secular liberals—then you see it as merely an act of love. Importantly, the harm of dyadic morality is intuitive and perceived, not a reasoned matter of fact (Schein & Gray, 2015).

**Element 1: Positive and negative associations and empirical equivalence**

The simplest relationship in theory mapping is a positive or negative association between constructs, represented by a line connecting them (for positive correlations) or a line with a dash through it (for negative correlations). In Figure 1, we see the size of cars is positively related to safety in an accident but negatively related to fuel efficiency; compared with compact cars, trucks fare better in crashes but burn more gas.

In Figure 1, we also see that recognizing the vulnerability of the victim (i.e., patient) is tied to feelings of empathy. We empathize more with the vulnerable (e.g., children and animals) than with the less vulnerable (e.g., adult humans; Preston & de Waal, 2001), and the more we empathize with someone, the more vulnerable he or she seems (Dijker, 2010). Figure 1 also illustrates that empathy is negatively linked to emotion regulation. As empathy can be an aversive experience (Batson, Duncan, Ackerman, Buckley, & Birch, 1981), emotion regulation prompts us to suppress this feeling (Cameron & Payne, 2011).

In terms of validity, convergent validity is revealed by the presence of a line between relevant constructs (a significant correlation), whereas divergent validity is revealed
by no line between nonrelevant constructs (i.e., no significant correlation). For correlations of sufficient size (e.g., $r > .8$), one could argue that two constructs are empirically equivalent and therefore lack divergent validity—indicated by a line marked with $\approx$. For example, the immorality of popular “impurity” scenarios (e.g., having sex with a dead chicken; Graham et al., 2011) is predicted almost entirely by perceptions of harm ($r = .87$; Gray & Keeney, 2015), with the remaining variance accounted for by the weirdness of these bizarre scenarios.

**Element 2: Moderation**

Psychology has long emphasized the impact of moderating variables (e.g., individual differences) upon constructs. In theory mapping, moderators are in italics and enclosed within « ». Figure 2 shows that differences in a car’s horsepower predict its price (more powerful cars are more expensive).

Figure 2 also shows that psychopaths are less likely to feel empathy (Blair, 2005; Gray, Jenkins, Heberlein, & Wegner, 2011) and that those higher in autism are less likely to perceive intention (Baron-Cohen, Leslie, & Frith, 1985).

**Element 3: Fundamental elements**

Psychological experiences emerge from the combination of other, more fundamental elements (Lindquist, 2013). Just as you need the ingredients of butter, flour, and sugar to make different kinds of cookies, so too do you need a set of psychological ingredients to make psychological phenomena. For example, proper facial recognition needs the more basic processes of vision and memory (Calder & Young, 2005; Young, Hugenberg, Bernstein, & Sacco, 2012). In theory mapping, fundamental elements—those necessary to “construct” a phenomenon—are enclosed within an upward-pointing { symbol.

As Figure 3 displays, cars are constructed from the fundamental elements of an engine, a chassis, and a body. Figure 3 also outlines that the strength of moral judgment hinges upon the combination of norm violations, negative affect, and perceived harm. People morally condemn acts that break norms and make them feel negative (Nichols, 2004), but not all such acts are immoral (Turiel, Hildebrandt, Wainryb, & Saltzstein, 1991). For example, spitting in your soup at a restaurant is both gauche and gross but doesn’t warrant the same condemnation as spitting in someone else’s soup. The difference between these acts is perceived harm, as it seems like you could get sick from someone else’s spit. Consistent with this idea, perceptions of harm reliably predict the wrongness acts, even those that seem objectively “harmless” (Schein & Gray, 2015; Schein, Ritter, & Gray, 2016).

Harm itself is also made from more fundamental elements: an intentional agent, a vulnerable patient, and the causation of damage ($A \rightarrow P$; Gray, Waytz, & Young, 2012). This explains why a CEO kicking a little girl seems more harmful—and therefore more immoral—than a little girl throwing mud at a CEO: CEOs are seen as possessing more intention and less vulnerability than little girls (Wegner & Gray, 2016), and kicking causes more damage than throwing mud.

The “{” symbol allows theory mapping to span different levels of analysis. Neural processes can combine to make cognitive processes, which can combine to make social processes, which can combine to make small-group processes, which can combine to make cultural processes.
Element 4: Varieties or examples

Although psychological constructs are often discussed as single things, they clearly vary across time, person, and culture. In theory mapping, different varieties or examples are illustrated with a dotted line to grey text. In Figure 4, we can see the variety across car brands, as grouped by manufacturer nationality.

Variation in phenomena is facilitated by variation in the fundamental elements underlying a phenomenon. For example, different varieties of love emerge from different amounts of the underlying elements of intimacy, commitment, and passion (Sternberg, 1986). With baked goods, different varieties stem from different combinations of ingredients such as flour, butter, and eggs. With cars, different models are allowed by different engines and bodies.

Dyadic morality acknowledges variation in morality—moral pluralism (Shweder, 2012)—and suggests that it arises from different varieties of norms (i.e., values; Schwartz, 1999) combined with different varieties of perceived harm (Gray et al., 2012). These varieties of harm are provided by underlying flexibility in who or what is seen as perpetrators (agents), victims (patients), and causes of harm. For instance, when cultures moralize funeral practices, they do so because they see the immortal soul as vulnerable to harm (Shweder, 2012).

Figure 4 illustrates varieties of agents, patients, and kinds of causation. For example, people can see adults, corporations, and spirits as capable of intending harm; can see damage caused via physical assault, emotional abuse, vandalism, or sacrilege; and can see vulnerable patients in the social order, eternal souls, children, and the future self. Of

Fig. 3. Theory mapping displays how phenomena are constructed from the combination of more fundamental elements.

Fig. 4. Theory mapping displays how variability of a construct is provided by varieties of its fundamental elements.
course, not all varieties are equally salient/typical, and so the most salient/typical variety (or two) is underlined. Typically, when we think of “immorality,” we imagine an adult human as the intentional agent, someone vulnerable as the patient (e.g., a child), and the damage as physical. Together, these typical elemental varieties yield canonical varieties of immoral harm—child abuse or murder.

Theory mappers may sometimes wonder whether a construct is a “fundamental element” or a “variety or example.” Fundamental elements are the necessary subprocesses or subcomponents needed give rise to a phenomenon, whereas varieties of elements represent taxonomic or content diversity, as revealed by techniques such as factor analyses, cluster analyses, and anthropological descriptions.

**Element 5: Numbers and notes**

Numbers and notes help supplement details lost when transforming multidimensional knowledge to two-dimensional space. For example, there is a “1” next to “Children” in Figure 4, which references the idea that people always highlight children in moral rhetoric (“Think of the children”; Bryant, 1977; Schein, Goranson, & Gray, 2015). Notes can also specify operationalizations and provide justification for a potentially controversial connection.

**The theory map: Putting it all together**

These five elements can all be combined to yield an overall Theory Map. Although seemingly complex, theory maps are built only out of the simple elements reviewed above, which seek to maximize the information to “ink” ratio (Tufte, 1983). Theory maps can capture information about an entire program of research without pages of text, providing comprehensiveness, concreteness, and conciseness.

**Cars.** A Theory Map for cars is displayed in Figure 5, revealing (a) the construction of cars from chassis, engine, and body; (b) the construction of the engine from crank shaft, cylinder block, and spark plugs; (c) varieties (brands) of cars; (d) moderators and their link to features such as price and fuel efficiency; and (e) varieties in engine and body type.

**Moral judgment.** A Theory Map for moral judgment is displayed in Figure 6, revealing (a) the construction of immorality from norms, affect, and perceived harm; (b) the construction of harm from the dyadic elements of agent, causation, and patient; (c) the varieties of affect (anger/disgust) and norms (authority/purity); (d) the varieties of agents (adults/corporations), causation (abuse/vandalism), and patients (souls/social order); (e) related processes (empathy/emotion regulation); and (f) related moderating individual differences (RWA/autism/psychopathy).

**More theory maps.** To further illustrate theory mapping, several scholars graciously provided theory maps for their fields of expertise. Table 1 lists their names and topics, as well as a relevant publication for future reading. These maps—and all maps explored so far—are provided in the Supplementary Material available online.
All these theory maps are listed on www.theorymaps.org. This website is meant to be a resource for the field, containing information about the technique, map templates, and many maps. I will post any theory maps sent to me (after a brief informal review) so that they are available to all. Importantly, theory maps can be submitted by anyone—whether a famous researcher, a well-read undergraduate, or an entire graduate methods class working together. Maps can be submitted for any topic, including one that already has a map, as science thrives through a plurality of perspectives.

**Benefits of Theory Mapping**

Although individual pet Theories can entrench viewpoints and bias perceptions, mapping out theory helps transcend these limitations by providing specificity and synthesis.

**Specificity**

By formalizing associations between constructs, mapping brings specificity to theory, which is all too often left
vague (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986). This specificity may help resolve arguments by revealing that debates focus upon different levels of analysis or upon different associations. Of course, by revealing the larger context of debates, theory mapping can also reveal that superficial skirmishes are actually deep disagreements about the human mind—such as with different interpretations of emotion (Lindquist, Siegel, Quigley, & Barrett, 2013) and morality (Cameron, Lindquist, & Gray, 2015).

The specificity of theory mapping may help to prevent redundancies. How many times has an old construct been “discovered” under a new name because investigators failed to specify the broader set of connected phenomena? Theory mapping may also increase validity and reliability because the convergence and divergence of constructs can be evaluated with a glance and because connections between concepts can be concretely specified in advance.

**Synthesis**

Theory mapping helps evaluate both the coherence of grand ideas and the contribution of specific studies. By detailing the constructs involved in a research project, readers can easily see the theoretical consistency and scope of a paper. Authors get upset when editors reject papers for “insufficient theory,” and extensions of theory mapping could provide a way to quantify what has always been qualitative, whether through the sheer number of connections implied by a paper or through more complex network metrics (Rafols & Meyer, 2009).

By looking at the surrounding context, theory mapping also helps evaluate the likelihood of new findings. A phenomenon that is consistent with other nearby phenomena is more likely to replicate than one that contradicts them all. In other words, theory mapping helps establish Bayesian priors regarding the truth of an effect.

Theory can also move subfields past debates about whether a phenomenon is “true” versus “false.” Consider discussions about social priming—whether being subtly primed with words like “Florida” and “grey” makes people walk more slowly. Work in cognitive psychology reveals that concepts activate similar concepts (Collins & Loftus, 1975) and that cognition is embodied (Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005) but also that such primes do not automatically translate into behavior (Loersch & Payne, 2011). The interconnection of these findings provides a nuanced perspective upon social priming that goes beyond dichotomies and suggests that we need not discard a phenomenon even if its most counterintuitive demonstrations are controversial (for similar thoughts, see Dijksterhuis, 2014).

**Conclusion**

Good science requires reliable methods and rigorous theory, but there is little discussion about how to improve the rigor of theory. Theory mapping provides one way for connecting ideas, building knowledge structures, and making concrete what once was vague. This technique may prove helpful in journals, classrooms, and anywhere else we need to evaluate ideas and synthesize studies. In the quest to improve science, psychology has...
understandably focused upon reliability, but even if every study is perfect, we still need to build them into a grand structure of knowledge—and for that, some blueprints would be handy.

Acknowledgments
Thanks to all the scholars who contributed maps and gave feedback on how to improve theory mapping.

Declaration of Conflicting Interests
The author declared no conflicts of interest with respect to the authorship or the publication of this article.

Funding
This research was supported by the Templeton Funded Imagination Institute (RFP-15-11), National Science Foundation (SES-1534083), Russell Sage Foundation (93-16-08), and Charles Koch Foundation.

Supplemental Material
Additional supporting information may be found at http://journals.sagepub.com/doi/suppl/10.1177/1745691617691949

Notes
1. In this case, there is a line when |r| > .3
2. The idea that psychological phenomena emerge from a combination of more fundamental elements is a key tenet of constructionist theories (Barrett, 2013).

References
Cameron, C. D., Hutcheson, C., Ferguson, A., Scheffer, J., & Inzlicht, M. (2016). Empathy is a choice: People are empathy misers because they are cognitive misers. Manuscript under review.


