

## A Rediscovered Ancient History of Motivational Interviewing and its Measurement

Christopher W. Dunn, PhD<sup>1</sup>, Sarah Peregrine Lord, BS<sup>1</sup>, Jessica Lowe, BS<sup>1</sup>, Jutta Joesch, PhD<sup>1</sup>, David C. Atkins, PhD<sup>1</sup>

### Abstract

We wrote this piece for coding teams around the world, hoping to raise some measurement issues, to inspire, and to entertain (perhaps not in that order). This one's for you, coders and for you, trainers of coders, you who work so hard to measure Motivational Interviewing using the standardized coding systems such as the MISC, MITI, and SCOPE.

### Keywords

*motivational interviewing, coding*

### SETTING

We are transported to ancient Pompeii (79 CE), where a spirited dialogue ensues in the sunny town square, between Epistemopheles and Methodia, two collaborative colleagues in search of Truth.

### ACTORS

*Epistemopheles*, formerly a “rock star” among Greek philosophers, has recently moved to Pompeii, having been ostracized from Greece for his hyper-empirical epistemology (theory of knowledge). He is infamous for having once said, “If you can’t see it, touch it, or smell it, it ain’t there.” He wears a dusty philosopher’s robe.

*Methodia*, a social scientist who once spent a year in meditative seclusion developing “Motivational **Intraviewing**,” the ancient method for privately changing one’s own bad habits using self-centered listening and directive self-talk. She is eager to share with Epistemopheles and hear his ideas about a new turn she has recently taken in her work. She wears an ancient stethoscope for listening to her own heart.

*Statisticuss*, a wizened gentleman who mysteriously appeared one day in Pompeii calling himself “Doc.” The town renamed him “Statisticuss” for his absurd belief that complex interpersonal interactions could actually be measured. The Pompeii gossip was that he may have come from another time.

### DIALOGUE

#### EPISTEMOPHELES

How good it is to see you again, Methodia my friend! First, let me

<sup>1</sup> Department of Psychiatry and Behavioral Sciences, Harborview Medical Center, University of Washington

Special thanks to Terri Moyers, Douglas F. Zatzick, Craig Fields.

The authors have no conflicts of interest to report.

Correspondence concerning this article should be addressed to: Christopher W. Dunn, PhD. Email: [cdunn@uw.edu](mailto:cdunn@uw.edu)

express my deepest appreciation to you for having developed Motivational **Intraviewing**. Although we philosophers have long thought that bad habits result from false knowledge, until now nothing could be done about it! I recently used Motivational **Intraviewing** for my own obsessive pontificating. I found that listening to my own ambivalent self-talk was so annoying that I was able to significantly reduce this behavior!

#### METHODIA

Almost as if your self-reflection nourished that tiny pedantic caterpillar inside of you to emerge as a beautifully socialized butterfly!

#### EPISTEMOPHELES

Exactly! Wow, you really understand...

#### METHODIA

Well, thank you for affirming my efforts to develop this method, Epistemopheles. That certainly boosts my confidence! Anyway, I could never have developed Motivational **Intraviewing** without our collaborative dialogues! The way you acknowledged my abilities and autonomy to make up my own mind evoked much motivation within me. Indeed, it was the experience that two heads are better than one that has recently taken my work in a new direction.

#### EPISTEMOPHELES

*(Quiet and eliciting.)* Tell me a little bit more about that.

#### METHODIA

Wow! Something about the way you said that just really frees me up to tell you more. Hmmm...

*(Pulled into a soft-focus flashback.)*

One day, while alone in my lab developing the intra-method, I suddenly realized that all that self-focus had left me feeling incredibly isolated... always staring inward. I felt lonely and incomplete...yearning for change but unable to find my pulse of readiness... One day I realized that I was going insane from all that time alone, so I finally went outside and began talking to people in the town square about my bad habit of isolating. I noticed that some of those people listened to me in a manner that made me believe that I could change, but others had a manner of being with me that pushed me away from change and made me feel helpless and hopeless.

*(Emerging from flashback.)*

So, my friend, I have discovered that certain listeners can help someone find their own way of changing—a way of change that couldn't have been articulated in isolation!

#### EPISTEMOPHELES

*(Lightbulb moment.)* If I may be so bold as to complete your paragraph, it appears that you have discovered Motivational **Interviewing!**

#### METHODIA

Exactly! Hmm...Motivational **Interviewing?** Not the sexiest name for something so exciting.

#### EPISTEMOPHELES

Well, maybe you could call it “My MI” for short? It’s a snappy name, and you could get the credit?

#### METHODIA

I dunno, I’m trying to get away from that self-centered stuff. How about we just call it “MI” and then spice it up with some acronyms?

#### EPISTEMOPHELES

Darn. See, the ideas you come up with on your own are way better than any advice I could ever give you!

#### METHODIA

And yet, it is you who are collaborating with me and evoking these ideas; I am freed to think about things in ways I couldn't have found alone...

#### EPISTEMOPHELES

Hmmm....So far, you have observed that at least in your case, two heads are better than one, as long as one listens to another in a particular manner. As I think about it, you must be wondering what other bad habits MI might help with.

#### METHODIA

Great summary, Epi. And yes, I do indeed plan to test MI for high-risk ouzo consumption.

#### EPISTEMOPHELES

But if it seems to work for ouzo, how will you KNOW with any degree of certainty that the particular MI manner was indeed used by counselors and not some different manner of listening or speaking? In other words, how will you differentiate between manners?

#### METHODIA

Hmm. I suppose someone would have to listen to measure the counselors' talk!

#### STATISTICUSS

*(Jumping out from the nearby bushes wearing 3-D glasses and holding a test tube, calculator and rubber chicken.)*

Surprise! Greetings, ancient philosophers!

#### METHODIA and EPISTEMOPHELES

Gasp!

#### STATISTICUSS

I have been following your dialogue from the bushes and it seems Methodia has discovered that the *manner* in which one listens to others' ambivalence about their bad habits can influence whether that person changes! Indeed, your dialogue is uncannily similar to one that I was just spying on in Norway where students are evoking the same thing from a

kindly gentleman with a red beard. They hope to be able to measure some of this stuff too! The idea is simply to write the rules for MI and then score the counselors on how well they adhere to those rules.

#### METHODIA

Simple?

#### EPISTEMOPHELES

Norway?

#### STATISTICUSS

Oops, never mind...I must have jet lag or something. Anyway, during these many weeks of listening to different dialogues, I have worked out a way to measure MI so that it can be tested and taught.

#### EPISTEMOPHELES

Well, Statisticuss, first off, if it's okay with you I'd like to share that I'm a little creeped out that you have been watching us from the bushes all this time. On the other hand, your ideas are intriguing. I suppose one could write rules that define the elements and style of MI, a yeoman's task in and of itself. But assigning numbers to human discourse is akin to looking through a dense MISC. The SCOPE of potential disagreement and the many potential sources of error are MITI daunting!

#### STATISTICUSS

Well, you sound a little surprised by this whole thing. If you don't mind, I have a couple of ideas about a TOOL for measuring MI that I call a “rating system.” And the people using that tool to measure discourse are called “coders” because they will “code” utterances by speaker and listener to certain categories.

#### EPISTEMOPHELES

Harumpf. Call it anything you want, but you can't make a tool that measures an interpersonal interaction; nobody can measure TALK!

#### STATISTICUSS

Maybe, maybe not. But if we don't at least try to do so, how will the world know if someone is doing MI or not? So, with your permission, I would first like to describe how such a tool would be used, because form follows function.

#### EPISTEMOPHELES and METHODIA

Okay, sure.

“It is important to apply a standardized coding system for monitoring MI style and technique, so that skill level vs. outcome can be analyzed... Miller is currently developing a standardized coding system which was not used by any of the 29 studies under review, so we are less certain that the style of MI was captured by these studies than we are the techniques.”<sup>1</sup>

“It is imperative that an effort is made in future studies to describe precisely how motivational interviewing education is performed and how to use the methods in client counselling, allowing us all to learn more about how to increase and maximise its effect.”<sup>2</sup>

#### STATISTICUSS

The first thing an MI rating system must be able to do is to *differentiate* between what is truly MI and what is not MI. That's the most basic need for such a tool, because, for example, if you use an MI style of counseling with a risky ouzo drinker, and later they drink less ouzo, how

do you know that it was MI that worked and not something else? In other words, the first question to ask is, can MI be differentiated from other ways of counseling?

### EPISTEMOPHELES

Granted.

### STATISTICUSS

*(Musing.)* Hmm, something like Project MATCH might provide the opportunity to answer that one...

### METHODIA

Huh, Project MATCH?

Project MATCH included a motivational interviewing intervention and “treatments were discriminable in that therapists...rarely used techniques associated with comparison approaches.”<sup>3</sup>

### STATISTICUSS

Don't worry. All in time, all in time. So as I was saying, if you were to find that MI cannot be clearly differentiated from other forms of counseling, you would close up shop and stop developing your measuring tool, because you would have learned that MI isn't different enough from other styles or that MI cannot be detected by coders. But suppose you found that MI is indeed unique and easily differentiated from other forms of counseling. In that case, you must begin to study if MI works. And while you are doing that, you must develop a “gold standard” for MI...

Here's the deal: if the impact of MI is being tested with different bad habits and different types of people then every study ought to collect MI scores for each counselor. Once many scores have accumulated for many different counselors you will eventually know *how good counselors must be at MI to be effective. That will be your gold standard.* If you do not do this, you will soon be facing a conundrum in which hundreds of studies may have shown that MI works, but nobody will know which elements of MI were actually done nor how good those counselors were at performing those elements. That would get messy because the world will soon be clamoring to learn MI, but nobody will be able to tell them how well one must do MI in order to change bad habits.

“...few studies have detailed how interventionists were trained, provided documentation of the fidelity of delivery of MI, or included process measures to relate to outcomes.”<sup>4</sup>

There's no way around this first step. Early on, you must establish norms for people who perform MI, or it won't make much sense to ask how good someone is at MI, because there will be nothing to compare them to.

### METHODIA

Norms?

### STATISTICUSS

Yeah, norms are very cool numerical scores that tell you whether your MI skill is at, above, or below average compared to all other counselors. But norms also tell you what proportion of that normative sample are above and below your score. Very handy scores, those norms.

At the very least you should establish norms for the population of counselors in scientific studies (presumed to be highly skilled in MI) and

establish norms for the population of community counselors (assumed to be less skilled). That way, you could meaningfully evaluate how good a given counselor is compared to the norms of two other populations. Eventually, thresholds could be established to compare MI efficacy. That way, trainers wouldn't have to guess whether their work is good enough, or go back and try to establish those quality benchmarks after the fact.

### EPISTEMOPHELES

Well, Statisticuss, I would I agree that you *might* be able to design a rating system that differentiates MI from non-MI. But I am less confident that you will be able to design a rating system that can measure *whether one counselor is better at MI than another counselor.* That is a much tougher measurement task. I say this because the difference between the scores of a counselor doing MI and the MI scores of a counselor doing something else is probably very large, since after all, the counselor not doing MI isn't consciously trying to adhere to MI spirit or technique. But the difference in MI scores between two different counselors who are both trying to do MI is likely to be much smaller and therefore harder to detect with measurement. It's easier to differentiate a cow from a goat than one goat from another goat.

### STATISTICUSS

Agreed, Epistemopheles! It won't be easy. Nonetheless, I would suggest counting the occurrences certain MI behaviors. And because some things can't be captured by a behavior count I would throw in a few global rating scales to represent the overall style of the counselor. Things like your tone, or the way in which you be with the other person, you know, some measure of their, their...

### EPISTEMOPHELES

*(Excited.)* Their “be-ness?”

### METHODIA

You're going to measure their *WHAT?*

### STATISTICUSS

Trust me, I would call that one “spirit” so as to avoid future misunderstandings. We could define it as the way a listener is being when they foster a relationship that allows another person to think about and welcome change- a way of being collaborative, evocative, and supportive of autonomy.

### EPISTEMOPHELES

And you think you're gonna MEASURE all that stuff?

### METHODIA

Your second confront today! Oooh, let's put that one on the behavior list, because it will help to identify ways of being that are not MI consistent! Statisticuss, please say more about these coders. How many of them would I need?

### STATISTICUSS

It is usually better to have more than one coder, for logistical reasons. One advantage is that two coders can finish the job in half the time. Also, it can take about the same amount of time to train three or four coders as it does to train only one or two. More than one coder allows for a certain amount of friendly competition, dialogue and support while learning the coding system. Just as Methodia has discovered that two heads are better than one, the same is true when it comes to coders following the MI scoring rules to the greatest possible degree. Two people can talk their way to a better understanding of the coding rules than one coder can alone. Just make sure that they don't get too carried away when debating coding discrepancies. Remember, no “true scores” exist, so coders should hold the attachment to the codes lightly, use the coding

guidelines to resolve discrepancies and defer to the trainer when utterly confused. Coders should try not to overthink any utterance. When in doubt during a coding discrepancy, each coder should reference the coding guideline that he or she thinks applies, and then simply apply that rule as consistently as possible. Rather than obsess or ruminate, coders must learn to “guess n’ go.” Finally, trainers should intervene if coders become hostile during utterance discrepancy discussions (e.g., challenging each other to duels or lobbing water-balloons into adjacent cubicles).

### METHODIA

So having more than one coder is the best way of keeping each other honest and pushing each other to interpret coding rules as rigorously as possible.

### STATISTICUSS

Right. There is also a statistical advantage of having more than one coder. It is easier to establish intercoder reliability because the correlations between sets of scores will be stronger the more raters you have. It’s a mathematical thing. Speaking of math, you should probably use intraclass correlations to calculate intercoder agreement on the behavior counts, but don’t use them for calculating intercoder agreement on the global ratings. For the globals, use this absolute agreement rule that will one day become popular: on the 5-point global rating scale, any two raters should not differ by more than one point for 80% of their common ratings.

If the coders reach high agreement and interrater reliability, it is more likely that they have been able to consistently apply the coding rules across the coding team. If there are still too many inconsistencies--within each coder’s ratings or between coders then it is likely you will not have high intraclass correlations on behavior counts, or absolute agreement on the global ratings. So even though there can be some risk that they are agreeing in the wrong application of the coding rules, *the process of reaching agreement through expert-facilitated dialogue* is your best bet for removing what we call “measurement error.”

### METHODIA

So in other words 80% of all pairs of global ratings must be within one point of each other?

### STATISTICUSS

Yup. And one more thing about those intraclass correlations. Some scientists will tell you that they must be .6 or above for adequate intercoder agreement. But each of those correlations also has attached to it a confidence interval, and the 95% confidence interval shouldn’t be any larger than....

### EPISTEMOPHELES

Whoah! What is that rumbling I hear?

### METHODIA

Oh, that’s just Mt. Vesuvius, our very old, inactive volcano, don’t worry.

### STATISTICUSS

Despite one’s best efforts to be accurate, there is always some measurement err... is that the floor shaking?

### METHODIA

Vesuvius seems pretty feisty today.

### STATISTICUSS

That thing is gonna blow!!!

*(Statisticus jumps into his DeLorean with flux-capacitor and disappears with the knowledge of Motivational Interviewing and coding.)*

### EPISTEMOPHELES

Egads! A shiny horse zooming through the air! Wait, where did that Statisticuss go?

*(Volcano explodes.)*

### METHODIA

I could be wrong, Epistomopheles, but I fear that we are in danger. What are your thoughts about moving to safer ground?

### EPISTEMOPHELES

FORGET SPIRIT! RUN FOR IT!!!!

## EPILOGUE

### NARRATOR

Alas, dear reader, Epistomopheles, Methodia, and their fine work were buried forever beneath a blanket of fiery ash, never to be seen again.

We have learned from this rediscovered ancient history that psychotherapy coding is not without problems. Will your analysis be at the utterance level? Paragraph? Talk turn? Thematic level? How will you decide which speech units to measure? What if you fail to specify the best speech units for understanding MI’s effectiveness? After all, nobody before us has ever coded everything in MI to see what actually causes the change. And if you leave something behind, it will never get measured, noticed, or taught by MI trainers.

Psychotherapy coding is not without its problems. First, a coding system must choose only one speech unit for analysis, leaving behind all others: will your analysis be at the utterance level? Paragraph? Talk turn? Thematic level? And what if you fail to choose the best one for understanding MI? Another problem looms when one is forced to choose which codes to include in a coding system and which ones to abandon. Although one hopes to include only the ones that cause change, we don’t know which ones those are, because after all, nobody before us has ever coded everything in MI to see what actually causes the change. And if you leave something behind, it will never get measured, noticed, and then taught by MI trainers.

Finally, there is the issue of reductionism, whereby one pays a necessary price for the luxury of condensing or reducing many words (the MI interview) down to few words (the codes). This reduction necessarily risks losing the meaning, intent, or tenor of the encounter. Let us hope that Statisticuss, the gentleman with the red beard, and his colleagues can work something out...

## ENDNOTES

<sup>1</sup> Dunn, C., DeRoo, L., & Rivara, F. (2001). The use of brief interventions adapted from motivational interviewing across behavioral domains: A systematic review. *Addiction*, 96, 1725-1742

<sup>2</sup> Rubak, S. (2005). Motivational Interviewing: A systematic review and metaanalysis. *British Journal of General Practice*, 55, 305.

<sup>3</sup> Carroll, K., Connors, G., Cooney, N., DiClemente, C., Donovan, D., Kadden, R., Longabaugh, R., Rounsaville, B., Wirtz, P., & Zweben, A. (1998). Internal validity of Project MATCH treatments: Discriminability and integrity. *Journal of Consulting and Clinical Psychology*, 66, 290.

<sup>4</sup> Burke, B. L., Arkowitz, H., & Dunn, C. (2002). The efficacy of motivational interviewing. In W. R. Miller & S. Rollnick, *Motivational interviewing: Preparing people for change* (2<sup>nd</sup> ed.) (pp. 217-250). New York: Guilford.



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 United States License](https://creativecommons.org/licenses/by-nc-nd/3.0/).



This journal is published by the [University Library System](https://www.library.pitt.edu/) of the [University of Pittsburgh](https://www.pitt.edu/) as part of its [D-Scribe Digital Publishing Program](https://www.library.pitt.edu/dscribe/), and is cosponsored by the [Motivational Interviewing Network of Trainers](https://www.mitrip.org/).