

TEXT AS DATA

IN THE SOCIAL SCIENCES AND HUMANITIES

Roger Baumann

Sociology & Social Work / Peace and Justice Studies

Who am I as a researcher?

Roger Baumann joined the Hope College Department of Sociology and Social Work in 2020.

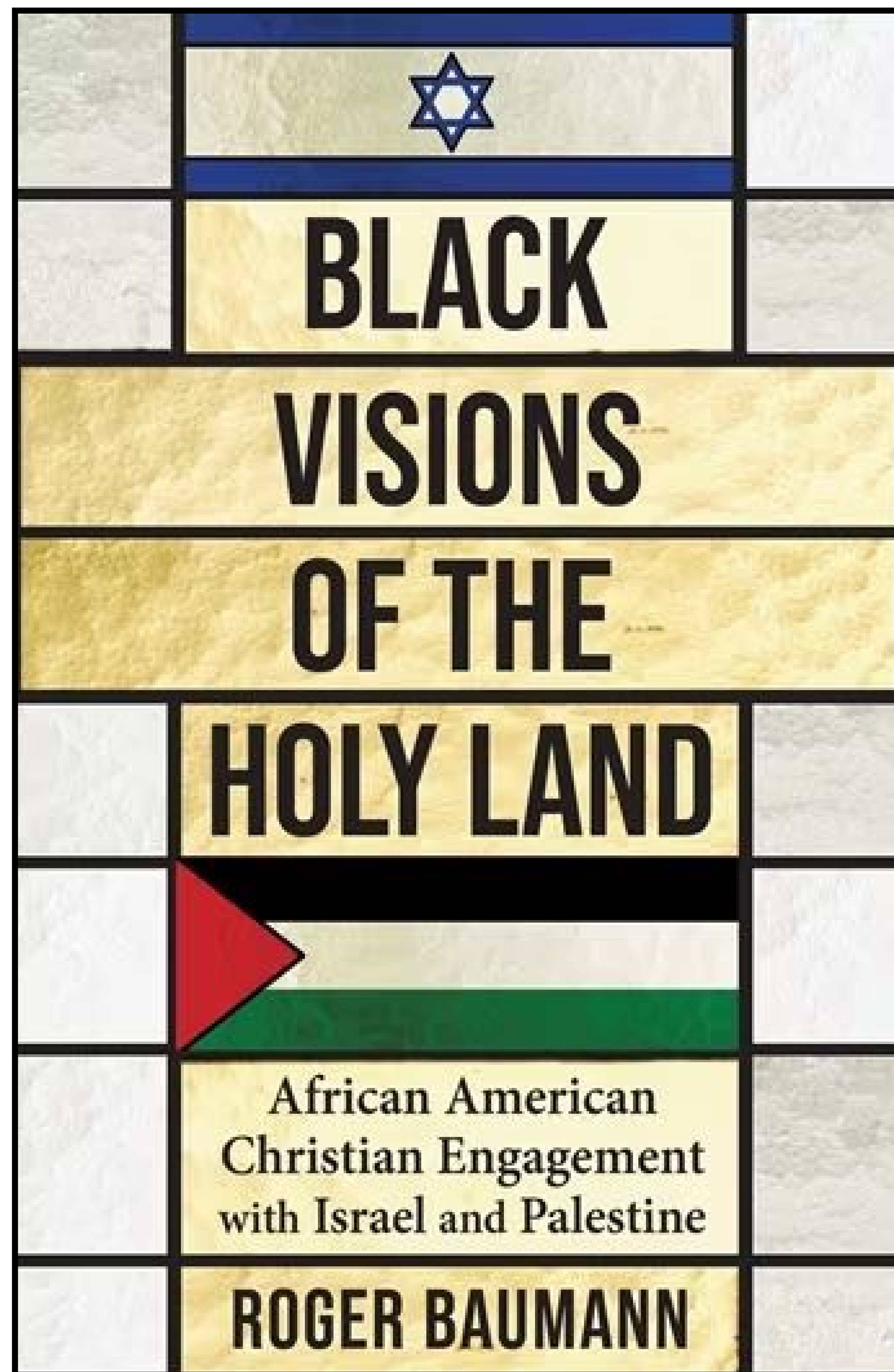
His research and teaching focus on questions about collective identity and social action. He is especially interested in how overlapping racial, religious and national identities contribute to how members of religious groups collectively understand who they are, where they come from and what they should be doing in the world.

His book — *Black Visions of the Holy Land* — analyzes African American Christian engagement with the global issue of Israel and Palestine. It examines how different groups of Black Christians invoke and contest the identity, history and mission of “the Black Church” as they approach that issue.

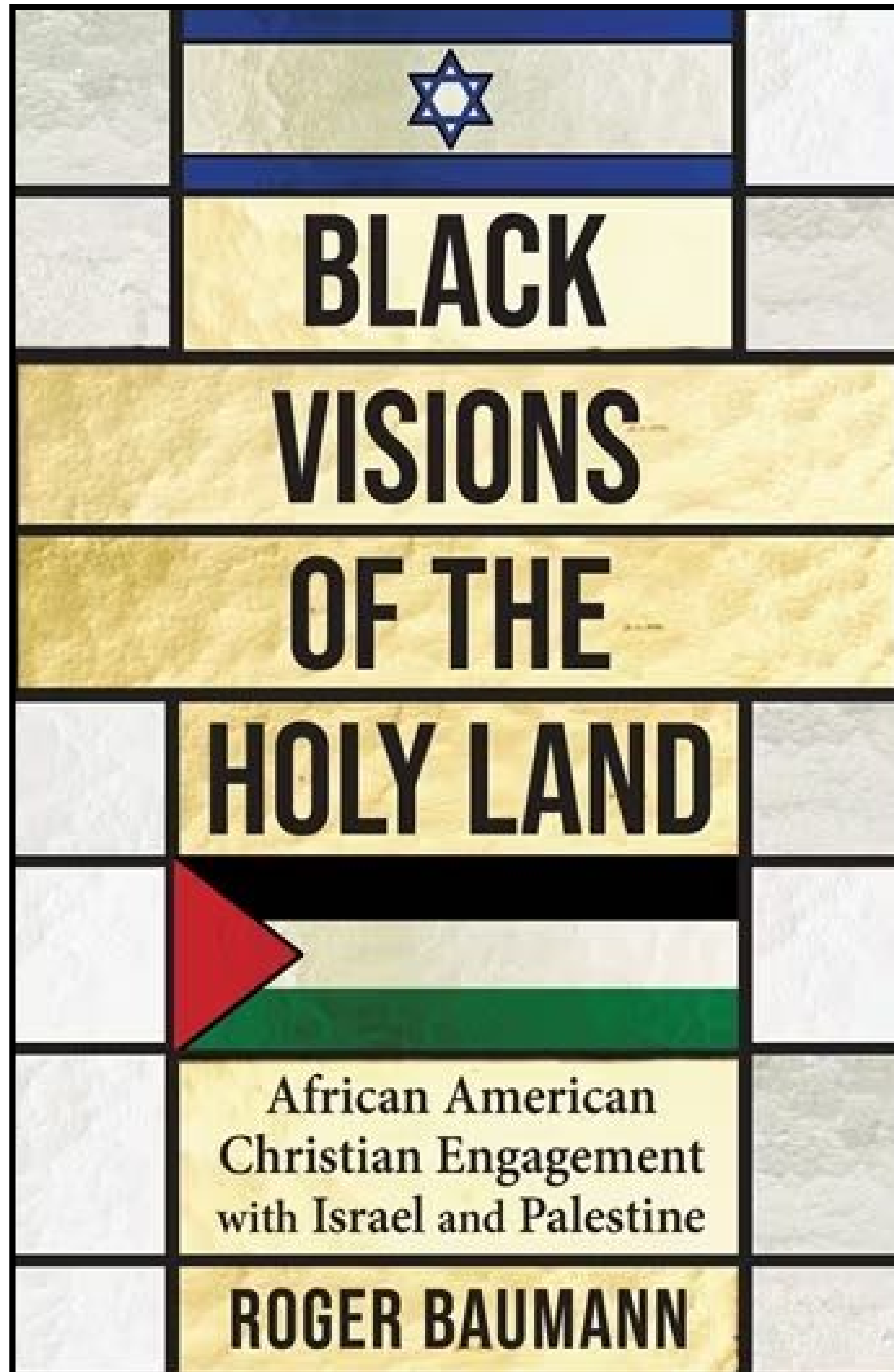
Another current project — *American Evangelicals, Islam & the Competition for Religious Authority* — analyzes hundreds of books written by and for evangelical Christians about Islam and Muslims in the last half century. It asks: What is the religious, political and social significance of American evangelical struggles over the authority to define Islam and Muslims? And what are the implications of this particular struggle for how evangelicals draw boundaries of inclusion and exclusion in the American public sphere?

Qualitative Research Methods





- What is “the Black Church”?
- Who gets to say?
- How does the meaning of race change from one social and political context to the next — including across national lines?



Data

- 90 Interviews
- Field Notes
- Analytical Memos
- Sermons and Speeches
- Social Media Posts
- Books, Newsletters, Emails

Making Sense of the Data







Home Import Codes Memos Variables Analysis Mixed Methods Visual Tools Reports MAXDictio Stats TeamCloud

New Project Open Project Reset Activations Logbook Teamwork Merge Projects Save Project As Save Anonymized Project As Project from Activated Documents External Files Archive Data

Documents Thanh (30 Paragraphs) 120%

Documents

- Interviews
 - Berlin
 - Mateo
 - Selim
 - Riley
 - Thanh
 - Chris
 - Hamburg
 - Focus Group
 - Video Interview
 - Images
 - Websites
 - Literature
 - Survey
 - YouTube: Etcoff: Happiness and its surpr...
- Sets
 - Respondents without children

857 438 228 36 70 46 41 35 210 84 4 4 2 55 184 86 438 195

on > negative > Sentiment > Slightly Negative

... ..
..Better use of free time
..Work/Life Separation
..negative
..Free time, self-care
..Workplace & Career

R: I would say I am not fully satisfied. I feel that right now, I'm not really using my free time efficiently. I would wish to use my free time more for my hobbies and to take better care of myself. Right now, with my free time, I'm a little lazy, so I watch Korean drama, or I just go to my neighbor's and then have a chat and drink tea, which is also nice, it's relaxing. But if I could plan my free time better, then I think I could. I would like to put some more self-care into my free time.

Q: How are your possibilities to plan your free time connected to your workplace situation?

R: I would say both work and free time are important. So, for work, I feel that when I'm satisfied with the work that I am doing, with the quality of the work I'm doing, the day goes by really quickly. And when I leave work, I don't feel tired, I just feel the same as when I started work. So, that's why I say both. I need to feel satisfied during the day and also after that I need to feel that I've done something for my personal life. I like to spend time with the people I love and do the things that I personally enjoy doing. Not doing it for money but doing it because I enjoy doing it.

Memo 38
I think it is remarkable that she works full time, takes care of a three-year-old, volunteers, and still talks about being lazy.

Codes

- Free time, self-care
- Work situation
- Wishes for the future
- Good quotes
- Satisfaction
 - negative
 - Sentiment
 - Positive
 - Slightly Positive
 - Neutral
 - Slightly Negative

20 138 40 7 0 47 0 2 9 22 9

Retrieved Segments

16 coded segments (from 1 document, 1 document group)

Interviews > Berlin > Thanh

Civil Status: Separated/divorced # of children: 1 Highest level of education: Master degree Employment: Employed Age: 32 Location: Berlin

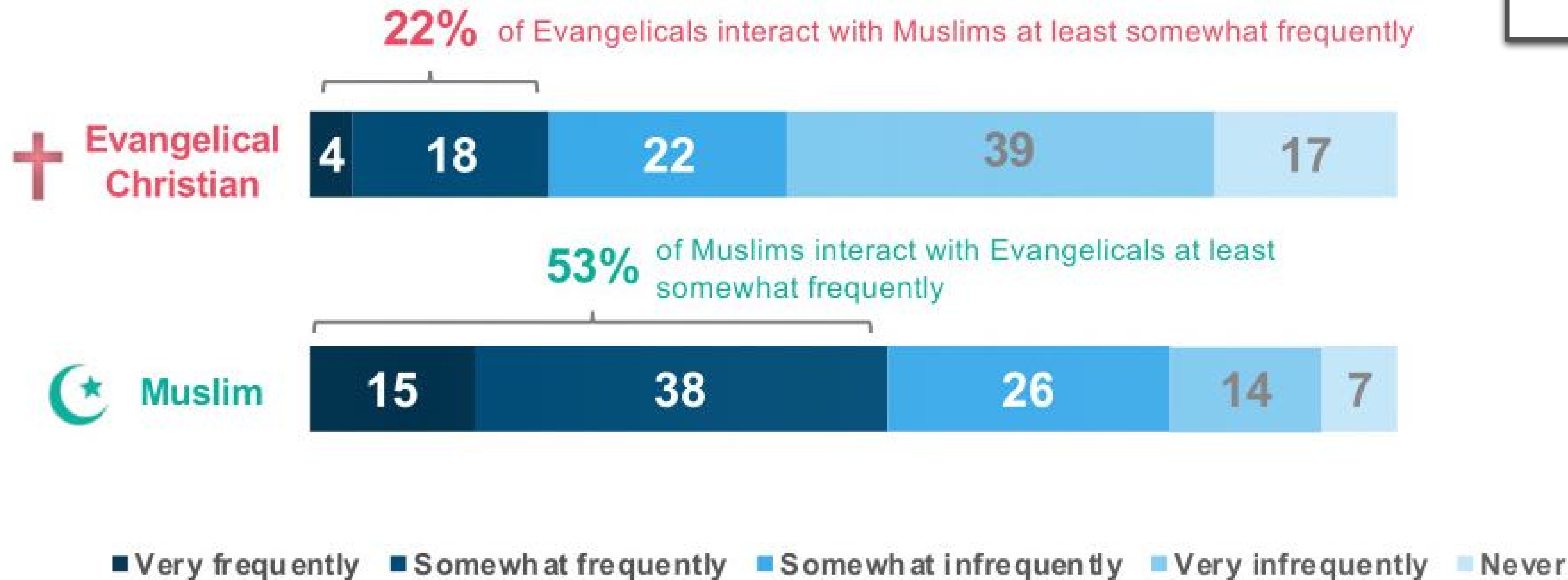
Okay. So I wake up at about 6:00 am and then I go out for jogging, two rounds around the block and I come back at about 6:40 am, and then I get ready for work. I am ready at about 7, 7:15 am. And then I eat breakfast. Sometimes, when I wake up half an hour later, I don't eat breakfast at home and I just kind of like buy something on the way to the office. So, and then I start working at about 8 to 8:30 am and get off work at about 5:00 pm, but then I go home. But it depends because I'm living separated from my ex-husband. And so half of the time. mv son is with me and half of the time mv son is with his dad. When mv son is not there. I just go home. And then I either go to a

Mode: Simple Coding Query

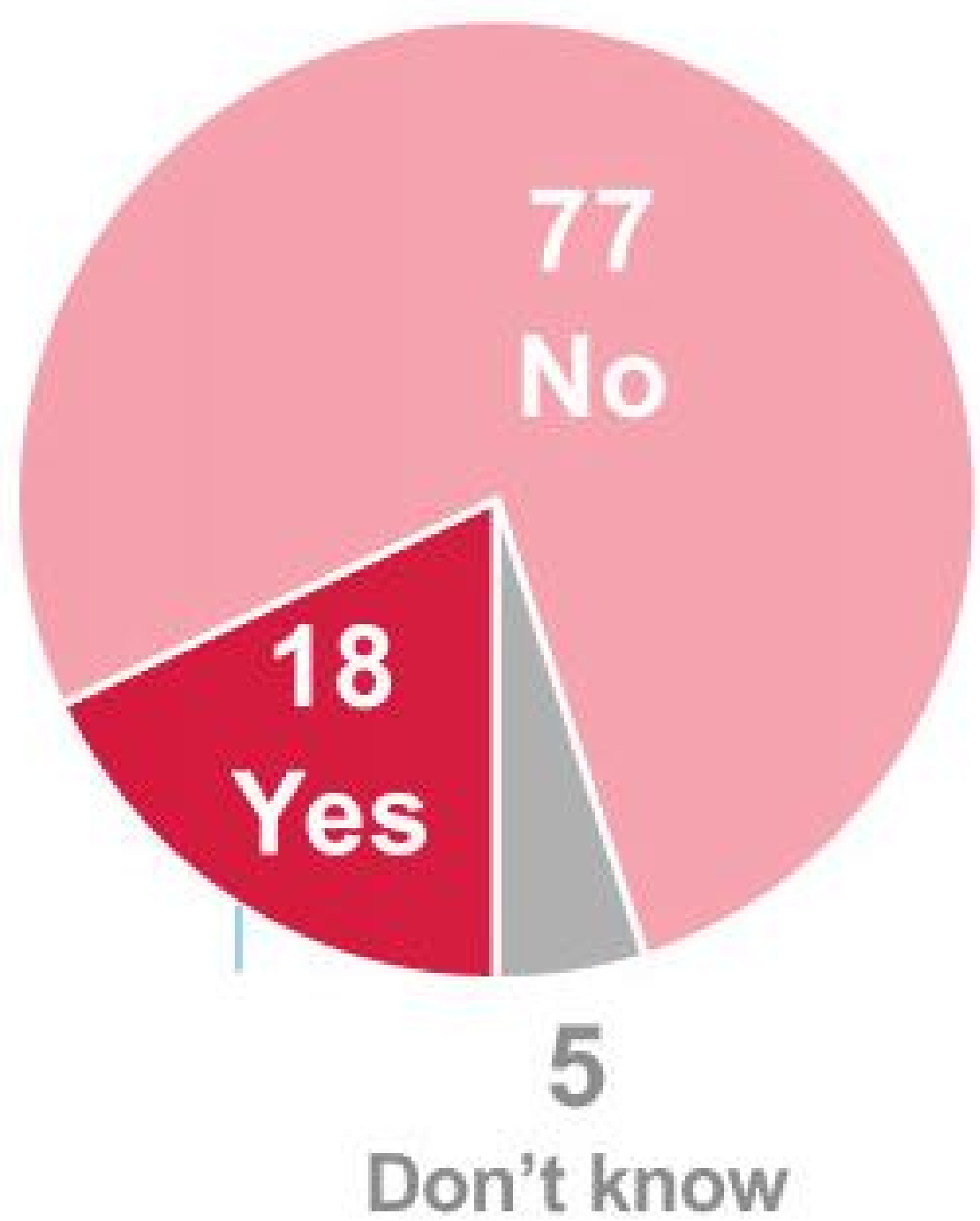


AMERICAN EVANGELICALS & ISLAM

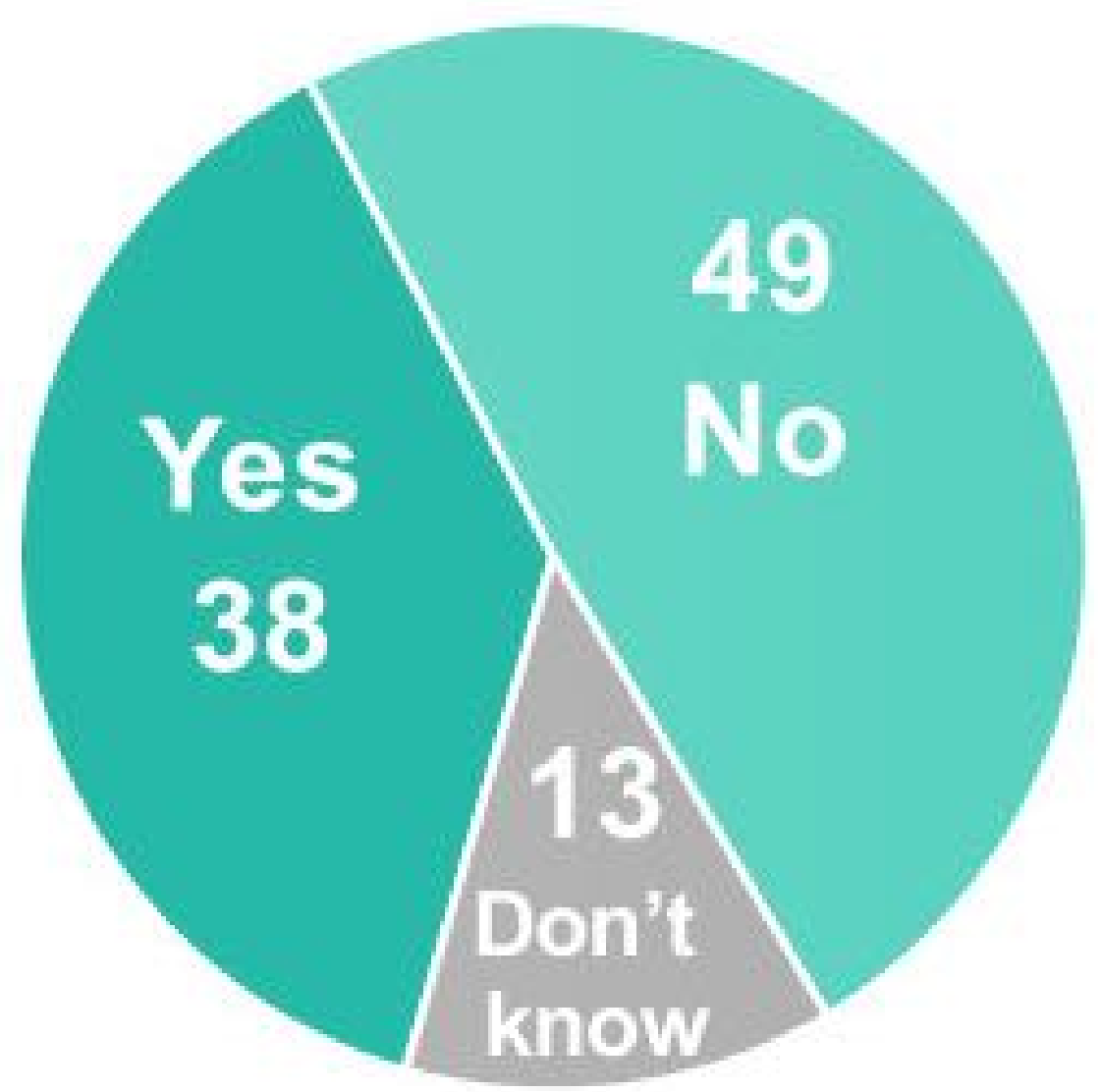
How often would you say you interact with people who are Muslim/Evangelical Christian?



Do you have any family members or close friends who are Muslim/Evangelical Christian?



✝ Evangelical Christian



☾ Muslim

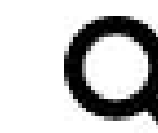


≡ SECTIONS

OUR LATEST

THE MAGAZINE

SUBSCRIBE



REVIEWS | BOOK REVIEW

Is Islam Receptive to Religious Freedom?

short-term outlook appears grim, says scholar Daniel Philpott, but there are grounds for hope.

MARSHALL | APRIL 15, 2019

INTERVIEW

Islam: Inherently Violent or Peaceful?

Scholar Warren Larson has a surprising answer to this difficult question.

STAN GUTHRIE | FEBRUARY 9, 2015

Shari'ah's Uphill Climb

Does Muslim law have a place in the American landscape?

VIEWS | INTERVIEW

Interview: Reopening Muslim Minds to Freedom and Tolerance

Mustafa Akyol explains why his new book explores the history of apostasy laws, the sayings of Muhammad, and the prospects for greater pluralism in the Islamic world.

TODD JOHNSON | JUNE 21, 2021

TRIBUTE

He Loved Muslims Because He Loved Jesus. The Bible Showed Him How.

Remembering the pilgrimage and legacy of Rick Love, who founded Peace Catalyst after years as international director of Frontiers.

JOSEPH CUMMING | AUGUST 18, 2020

TESTIMONY

I Laid Down My Islamic Privilege to Preach Jesus Around the World

How a direct descendant of Muhammad met Christ on a crowded Pakistani sidewalk.

CHRISTOPHER ALAM WITH PETER K. JOHNSON | APRIL 21, 2021

NEWS | RELIGIOUS FREEDOM

To Defend Mideast Christians, Can Advocates Critique Islam?

Diaspora leaders in America disagree on how to improve religious freedom back home.

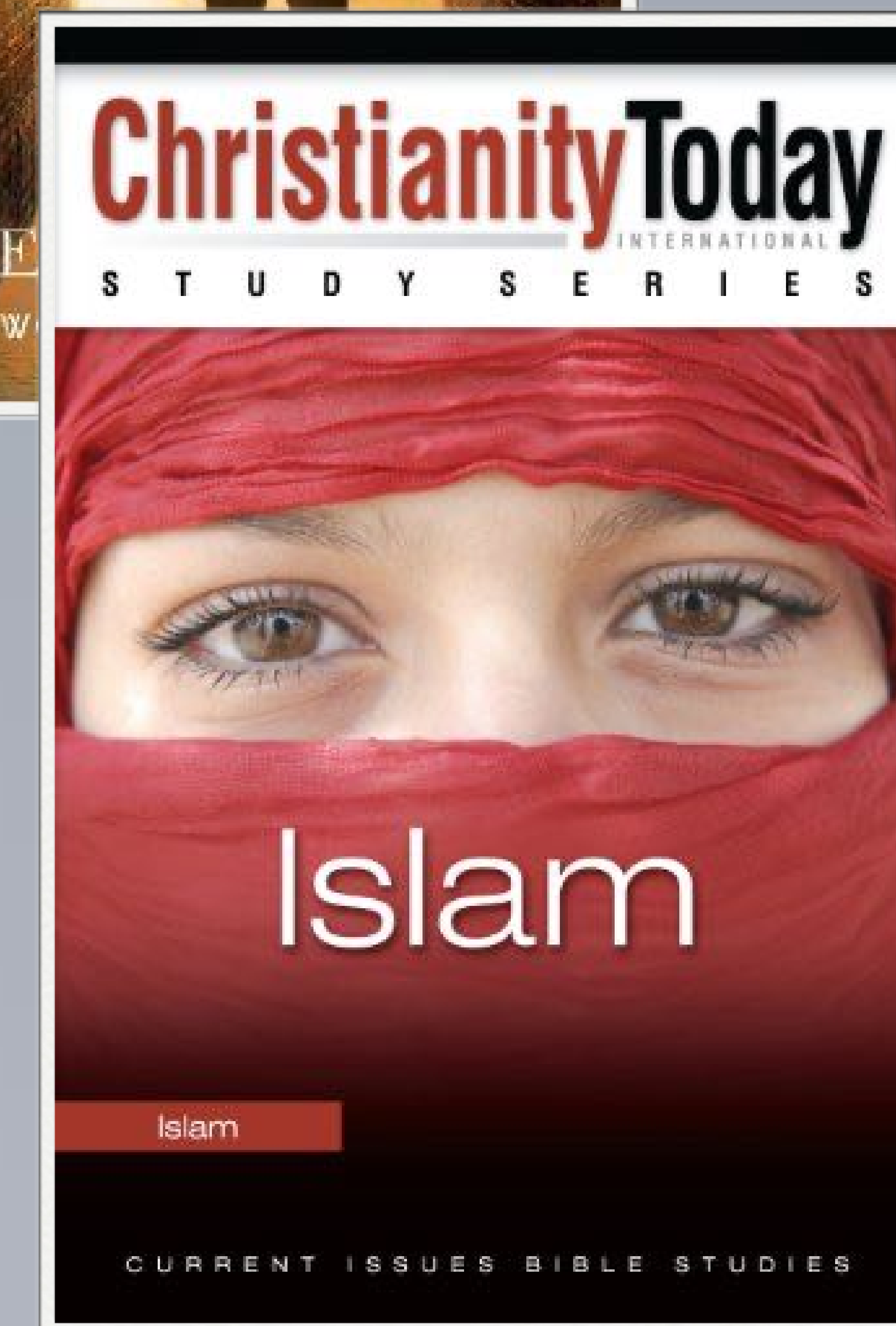
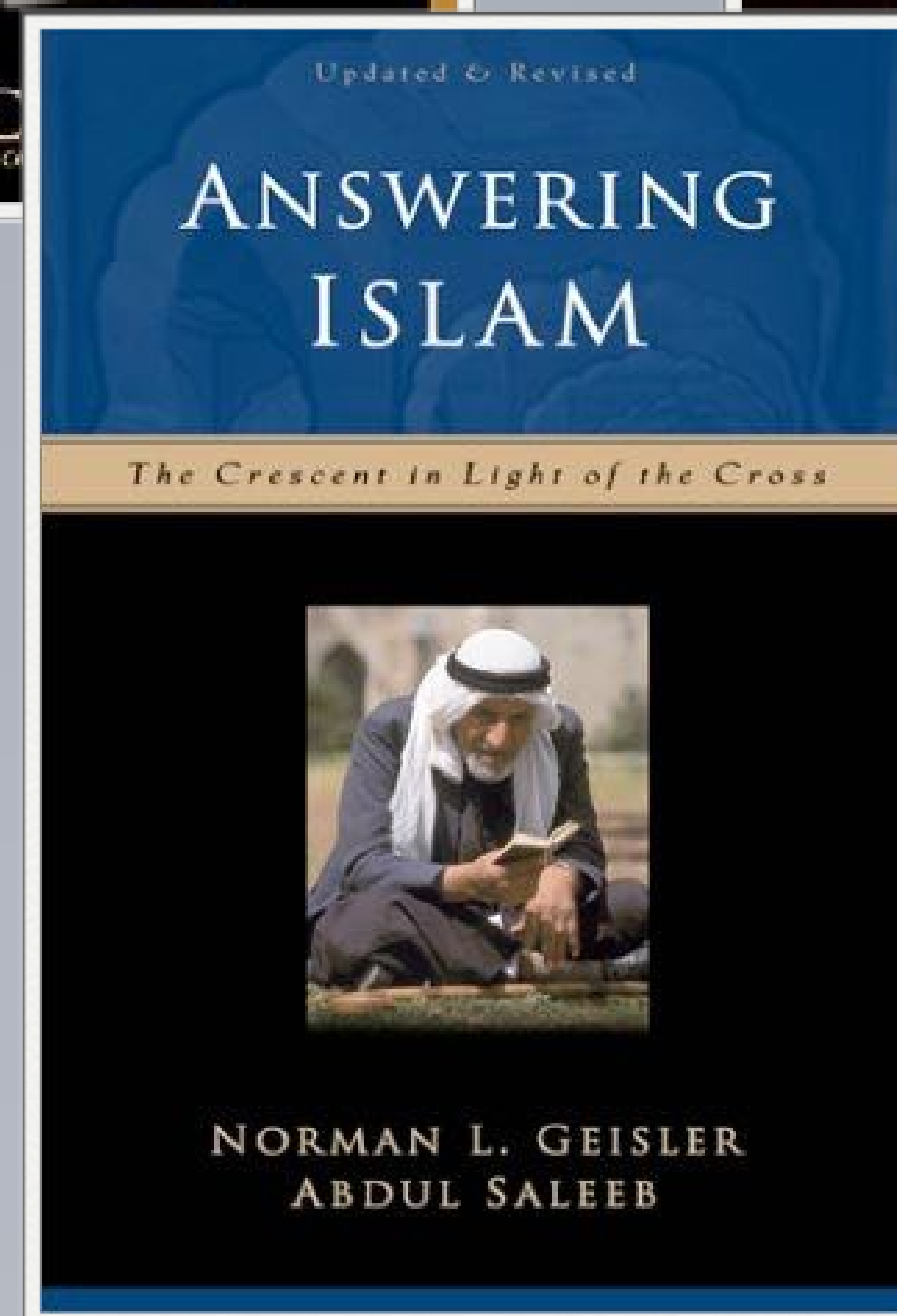
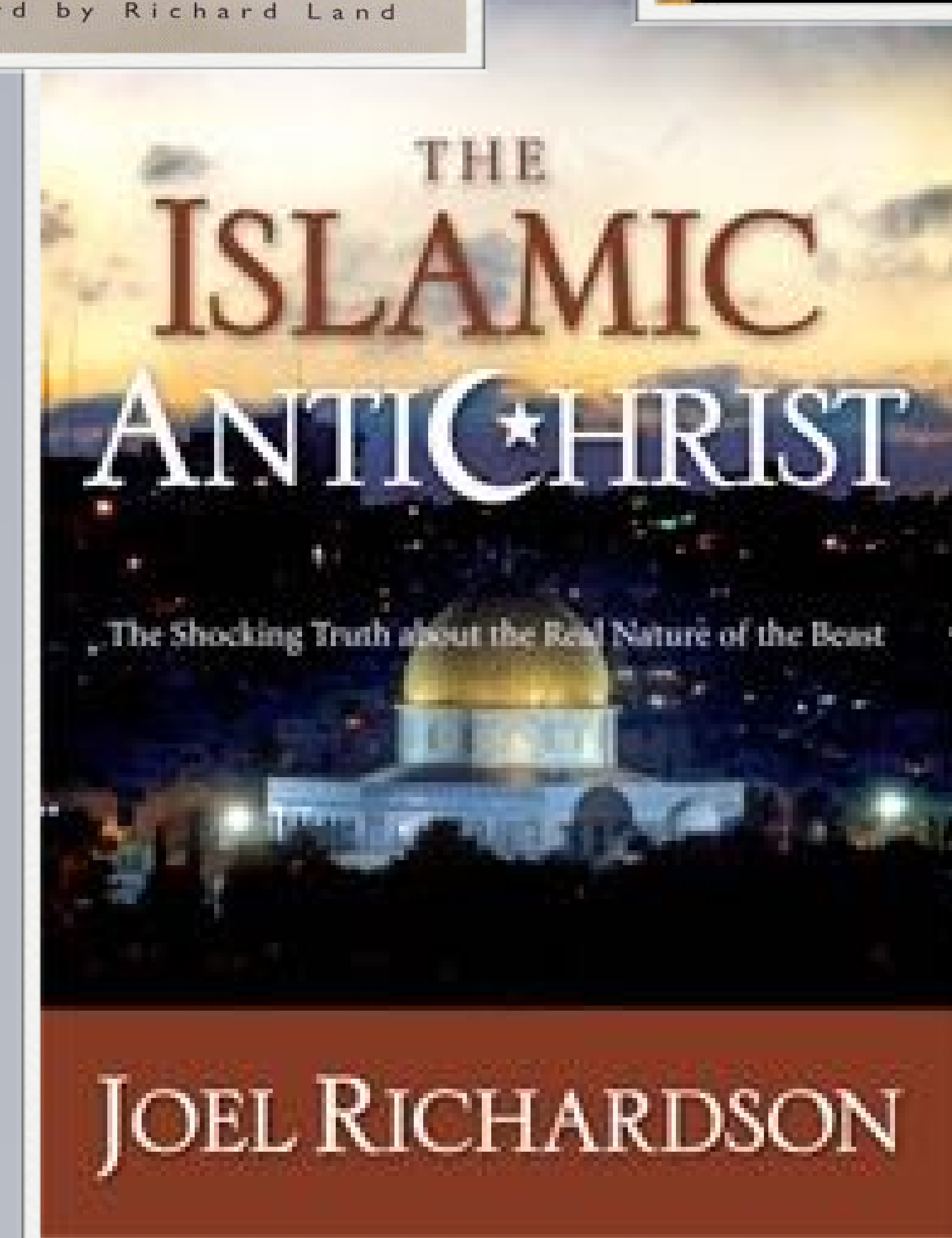
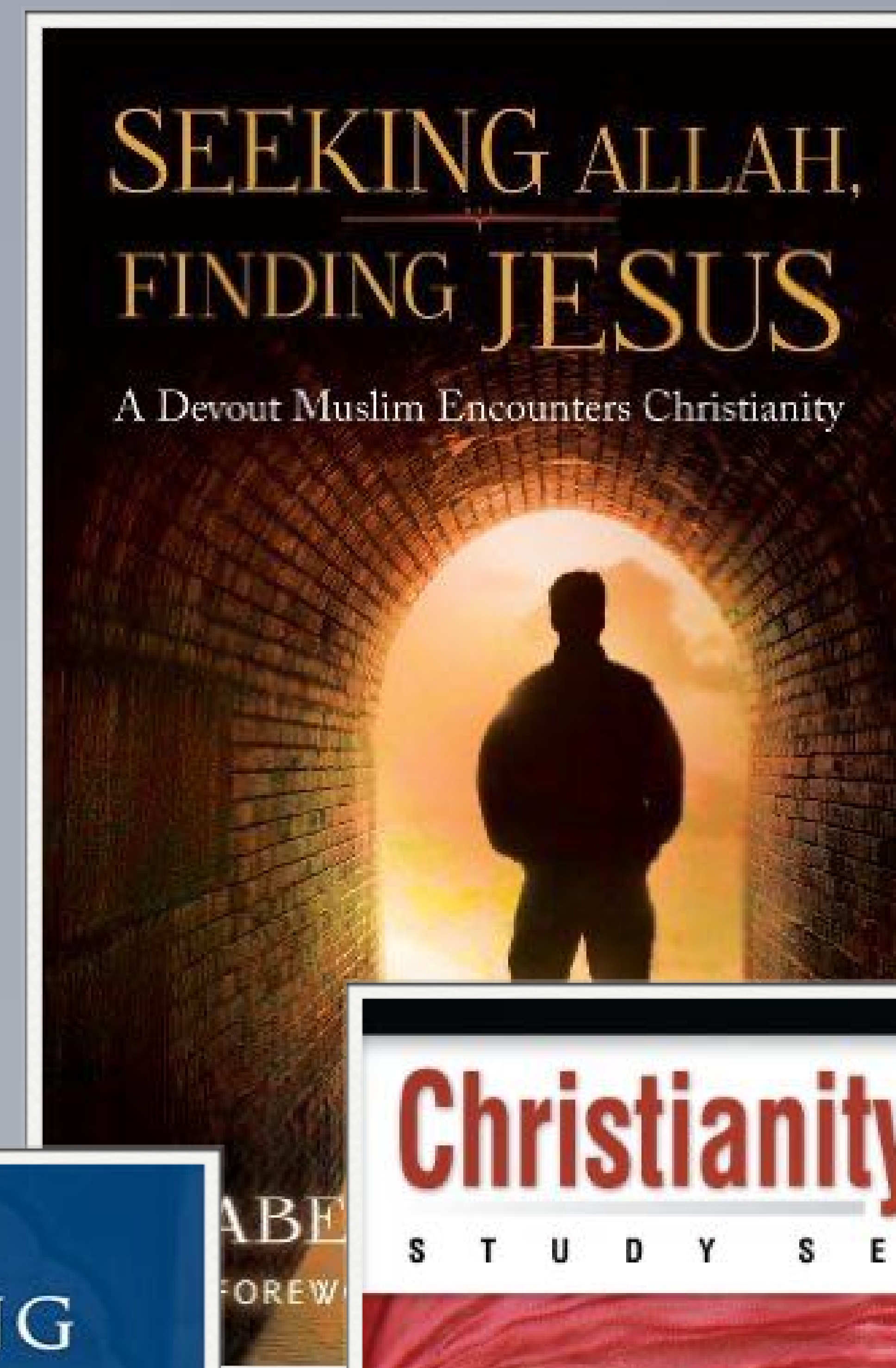
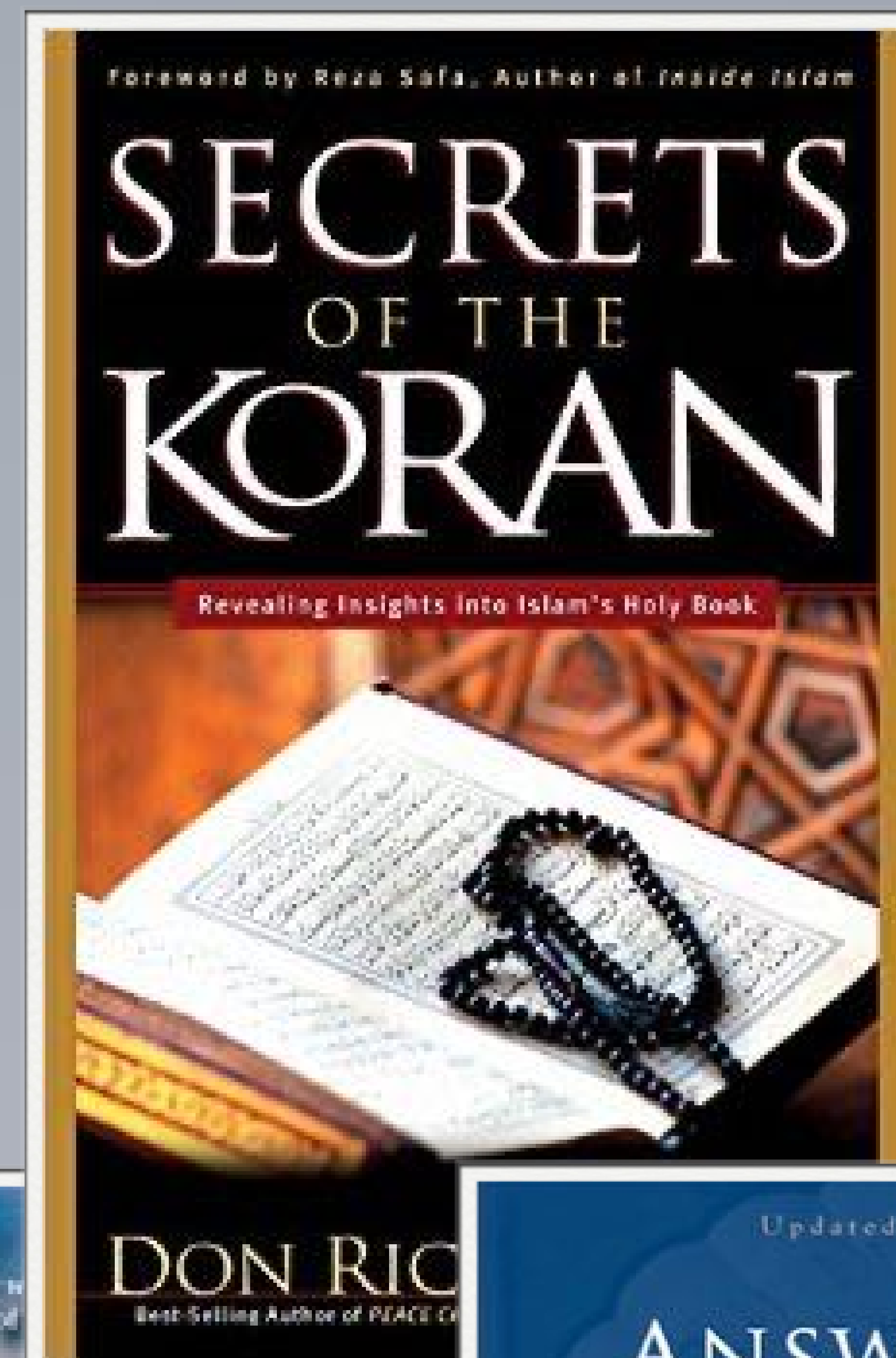
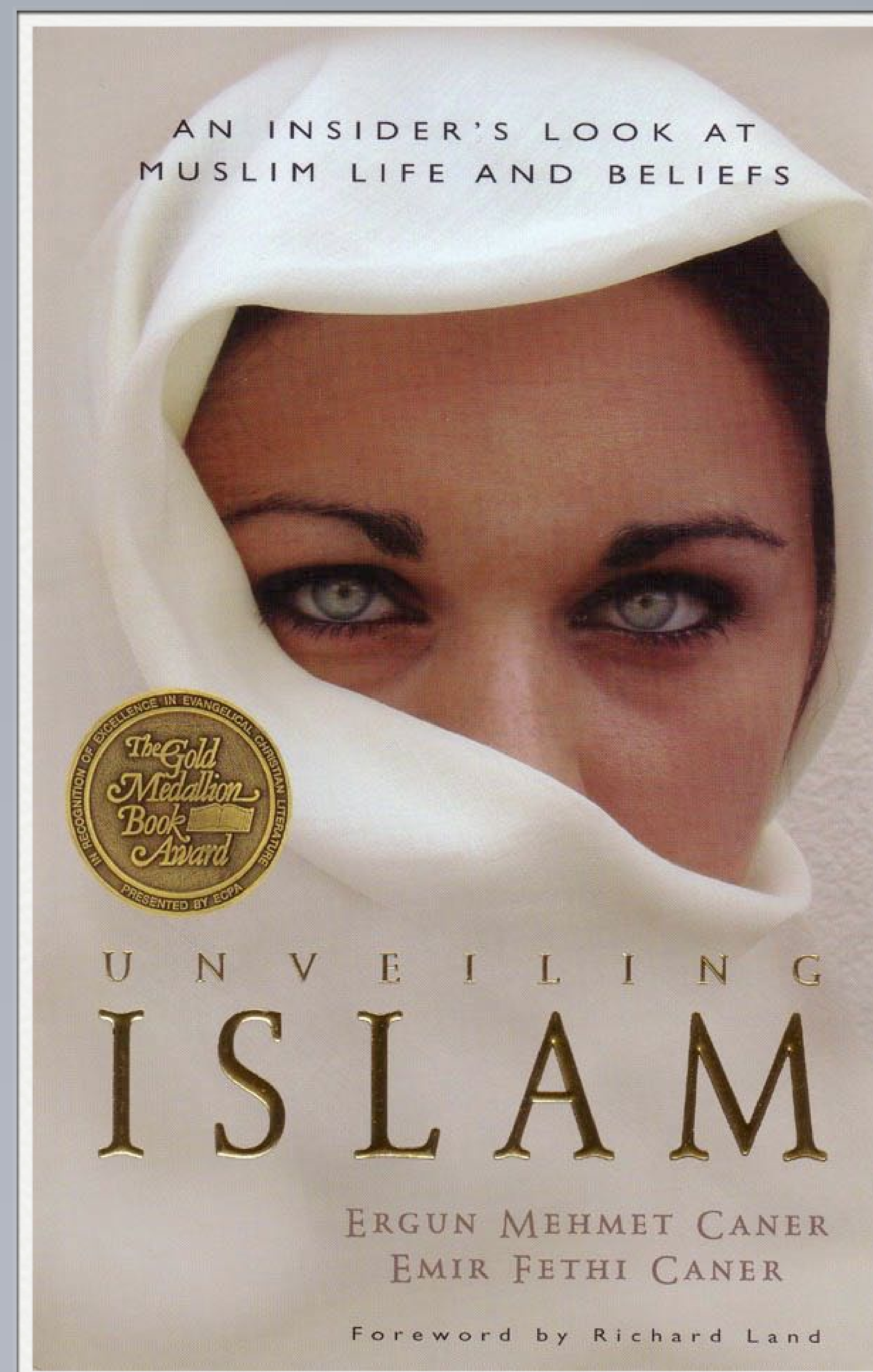
JAYSON CASPER IN CAIRO | DECEMBER 21, 2017

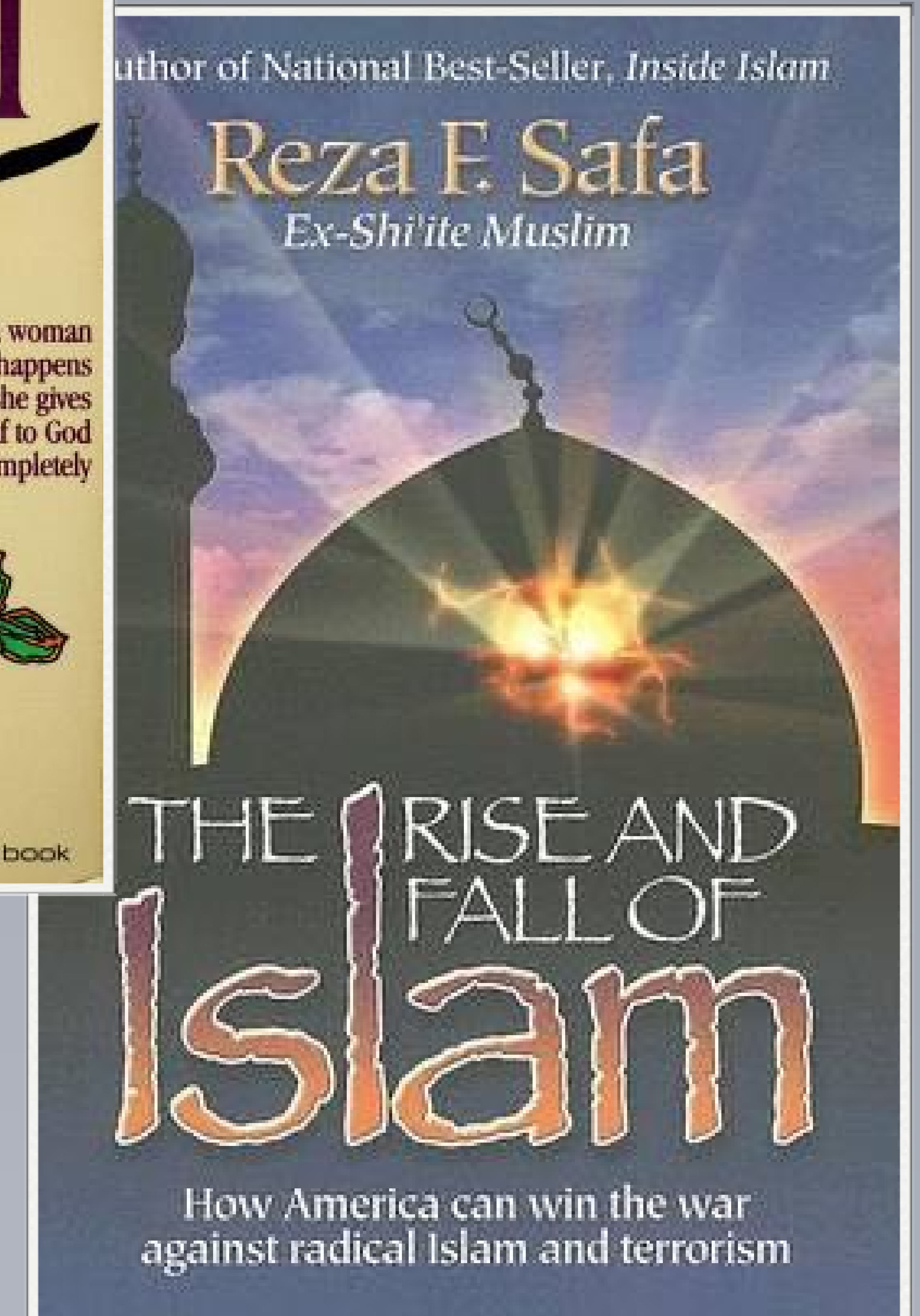
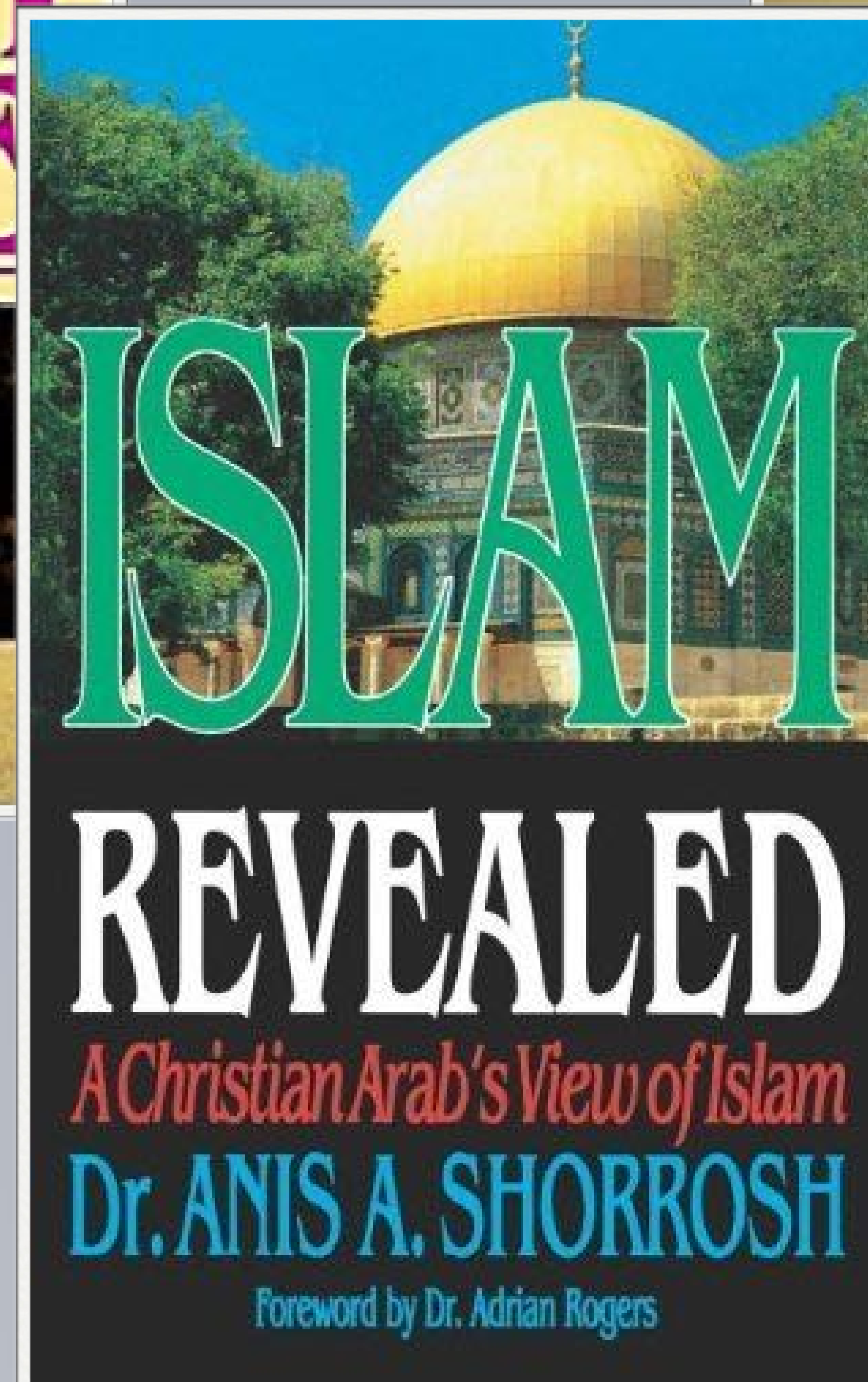
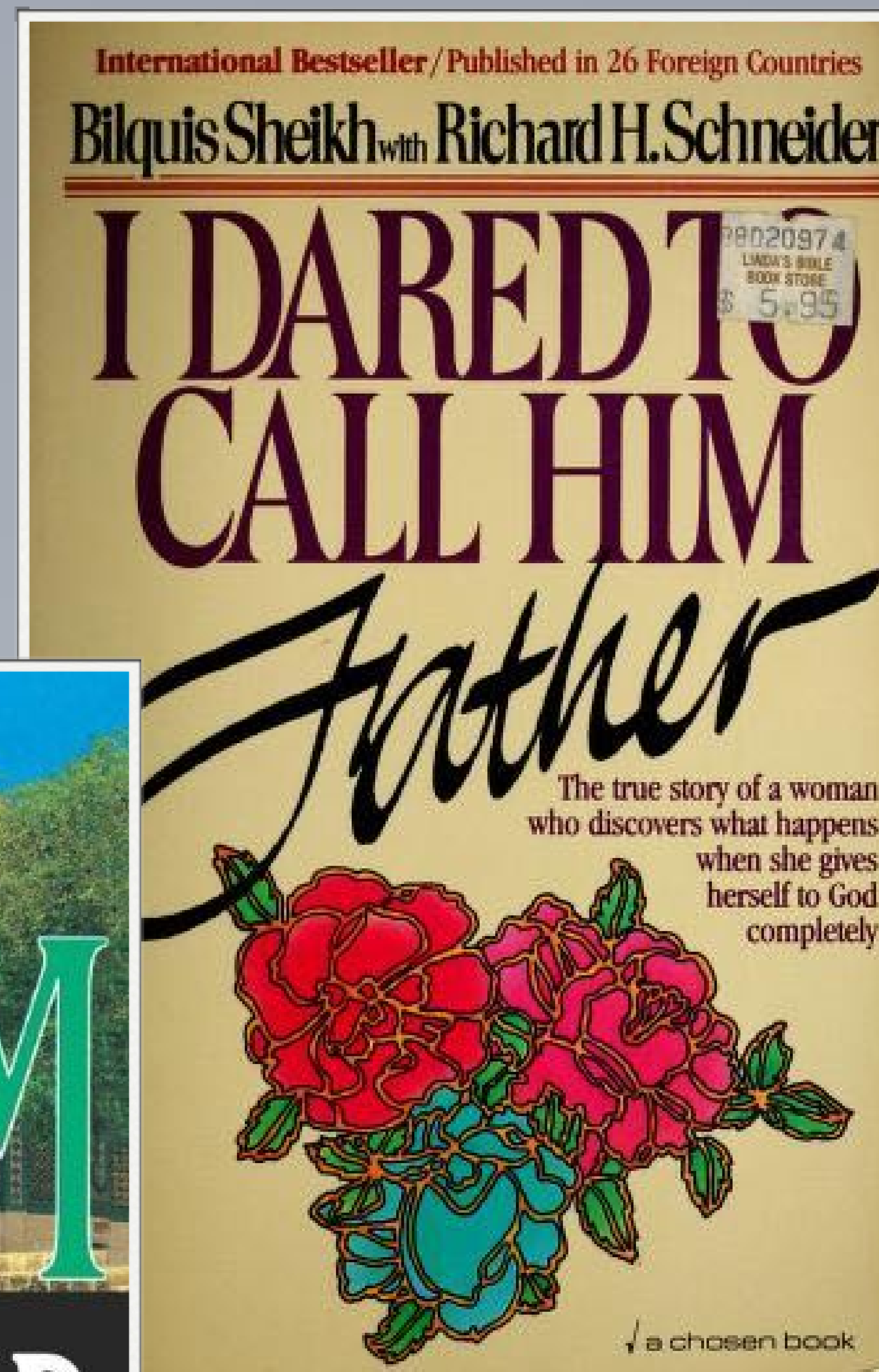
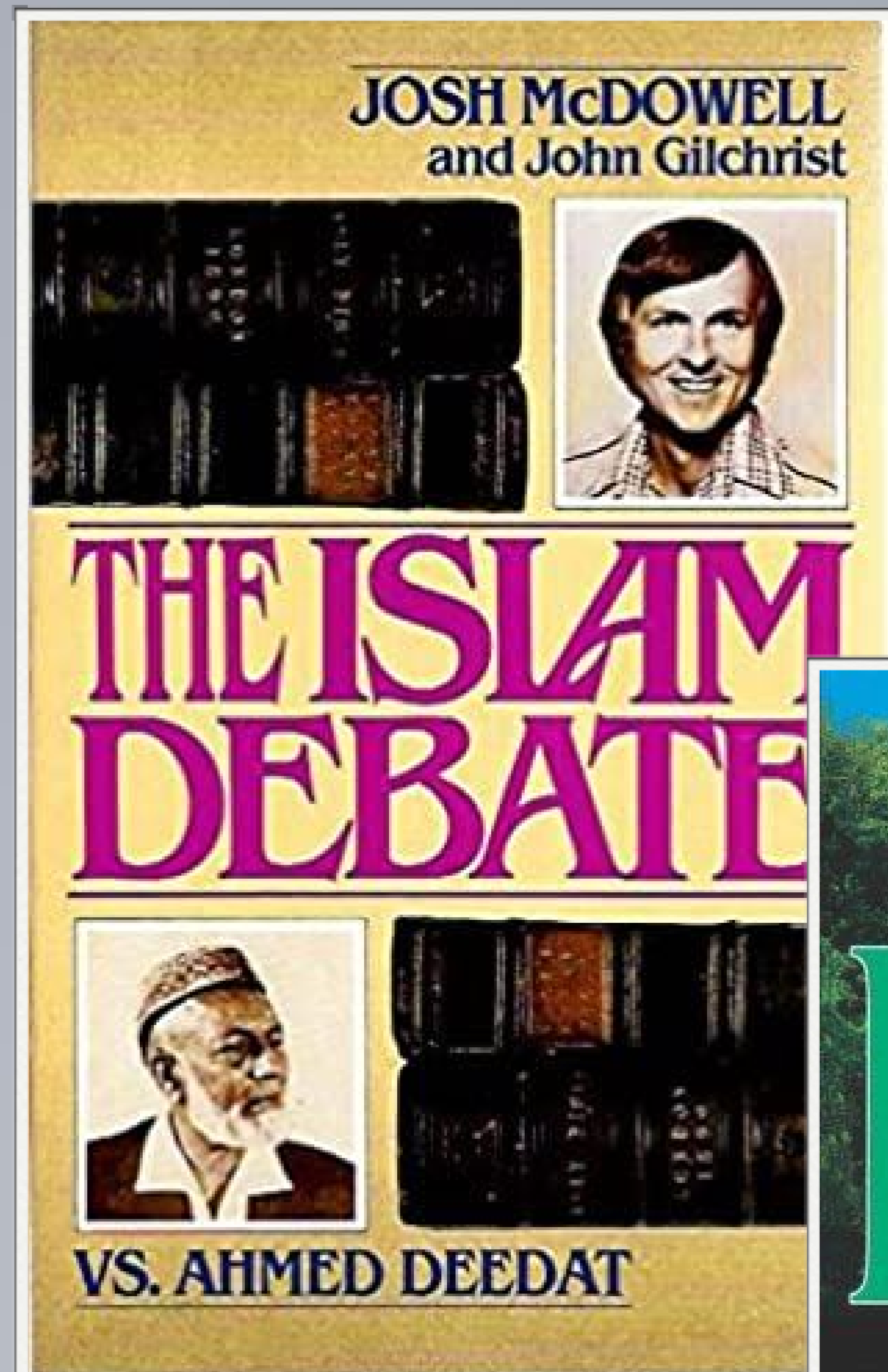
NEWS | CHURCH GROWTH

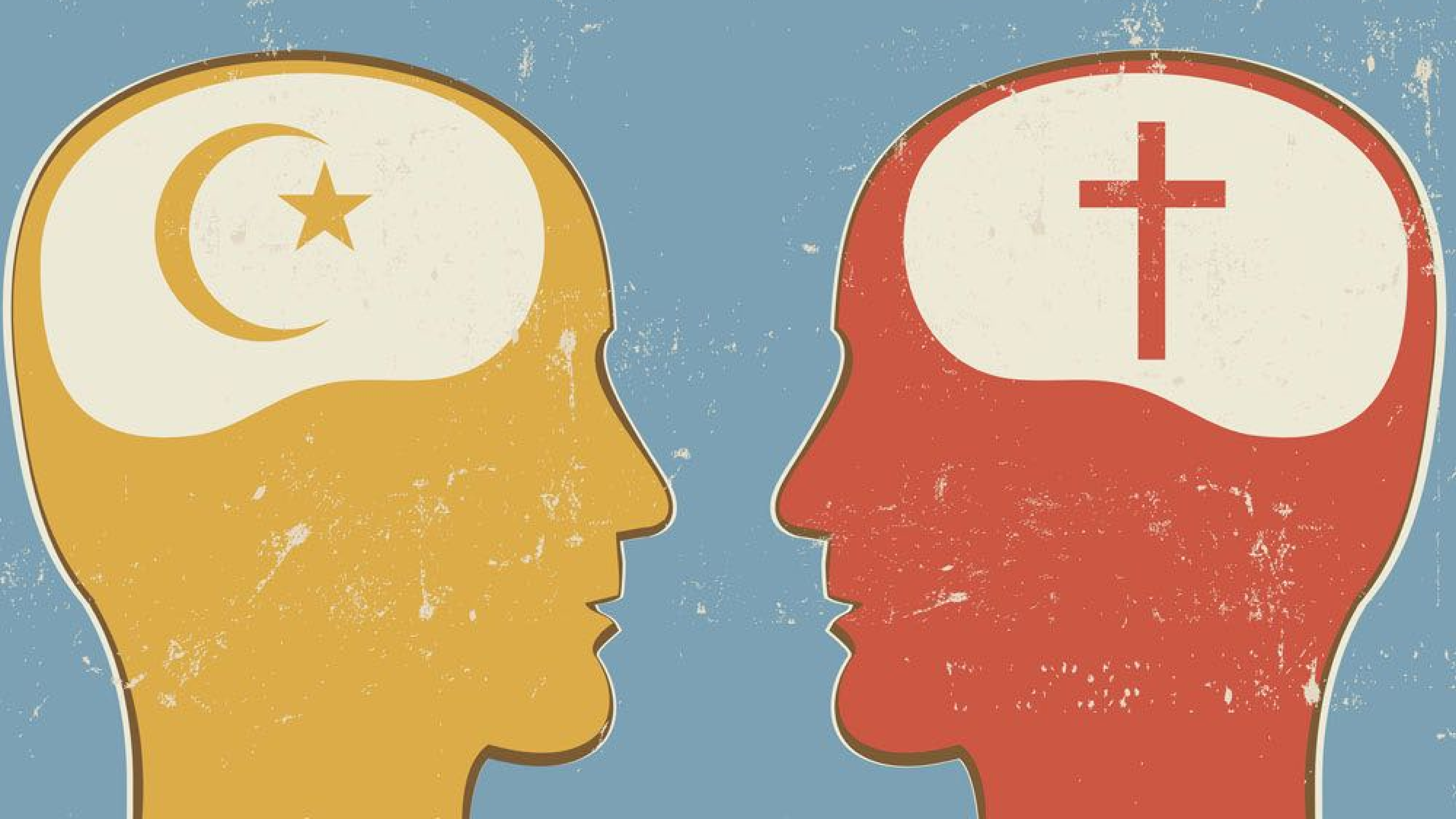
Why Christianity Is Surging in the Heart of Islam

Medical missions and market dynamics lead to millions of believers in the Arabian Peninsula.

JAYSON CASPER IN THE UNITED ARAB EMIRATES | SEPTEMBER 11, 2015

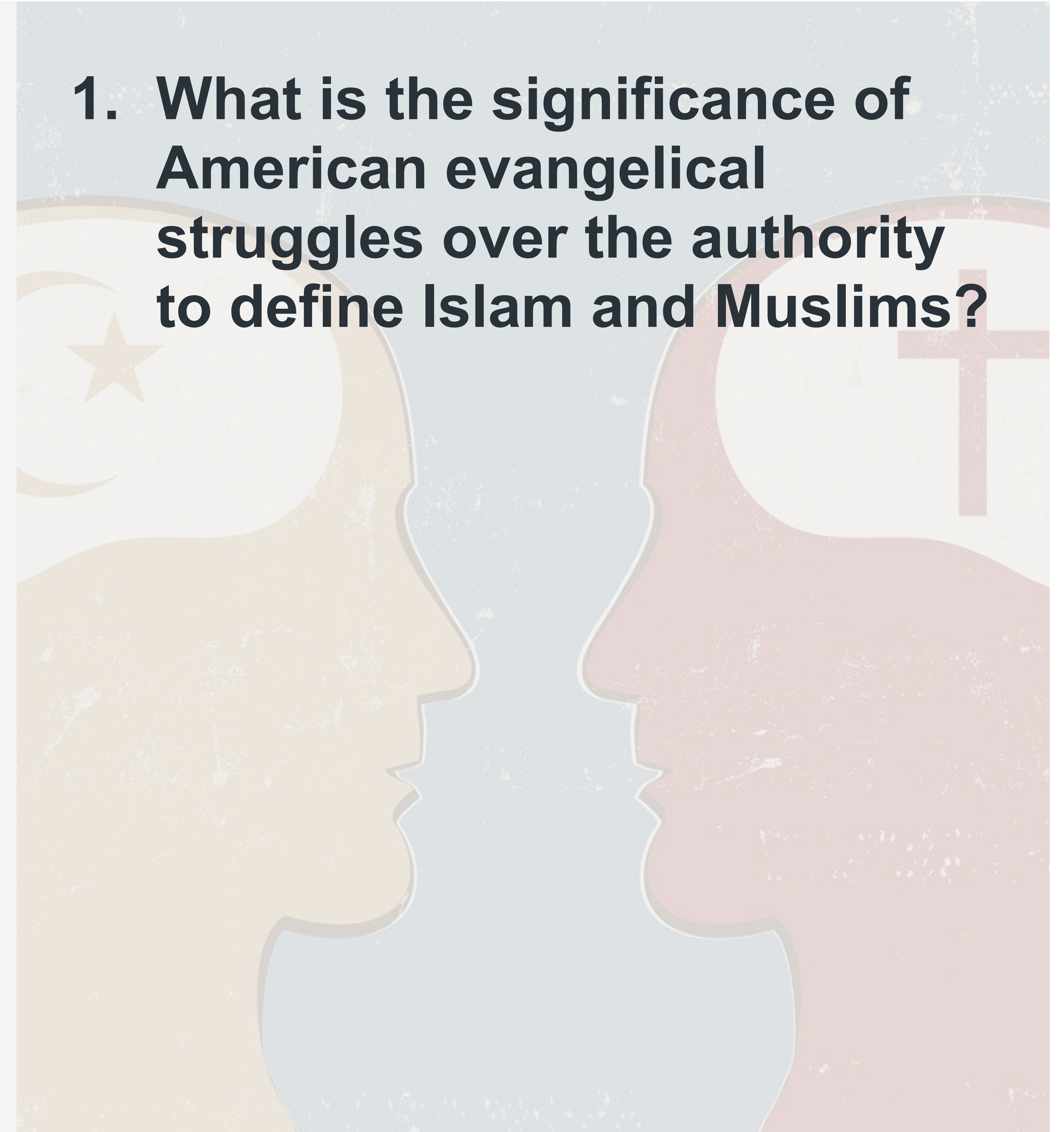






AMERICAN EVANGELICALS & ISLAM

- 1. What is the significance of American evangelical struggles over the authority to define Islam and Muslims?**



AMERICAN EVANGELICALS & ISLAM

- 1. What is the significance of American evangelical struggles over the authority to define Islam and Muslims?**
- 2. How do institutions, discourses, and practices about Islam and Muslims take shape?**

AMERICAN EVANGELICALS & ISLAM

- 1. What is the significance of American evangelical struggles over the authority to define Islam and Muslims?**
- 2. How do institutions, discourses, and practices about Islam and Muslims take shape?**
- 3. How do struggles to define and relate to Muslims contribute to enhancing or undermining civic engagement?**

AMERICAN EVANGELICALS & ISLAM





WHO ARE THE AUTHORS?

- **Apologists**
- **Academics**
- **Bible Prophecy Experts**
- **Activists**
- **Missionaries**
- **Writers/Journalists**
- **Pastors**
- **Evangelists**

WHAT IS THEIR POSITION VIS-A-VIS ISLAM?

- **Convert to Evangelical Christianity**
- **Middle Eastern Christian**
- **Other Experiential Knowledge (e.g., Missionaries, Military, Corporate)**
- **None of the Above?**

BOOK GENRES

- **Missiological**
- **Apologetic**
- **Apocalyptic**
- **Pastoral**
- **Testimonial**

THEMES ACROSS GENRES

- **Insider Knowledge**
- **Conversion**
- **Proselytizing**
- **Persecution**
- **Civilizational Threat**
- **Moral Threat**
- **Spiritual Warfare**
- **Jihad/Violence**
- **Peace**
- **Commonalities/Differences**
- **Dialogue/Cooperation**
- **Gender**
- **Sharia**
- **Israel**



AMONG THE NUMEROUS AT- TENDENCY TO BREAK AND C- FOR THEIR CHARACTER AND DUE VALUE ON ANY PLAN V- TICE, AND- IMENTS HA- DERIVE THE- BOTH AN- D THAT TH- ROM OUR- IT, THAT O- ES ARE TO- ITERESTED- E, OF KNO- SITUATION- IMENTS, BL- UND, PARTI- OED FROM- E WITH WI- N A MOUNT- VITAGES PROMISED BY J- TROL THE VIOLENCE OF- E, AS WHEN HE CONTRA- T, WITHOUT VIOLATING- I INTRODU- IERE PERIOD- CIOUS DEC- ODERN C- EFFECTUALLY OBIATED- DERATE AND VIRTUOU- AENTS ARE TOO UNSTA- IDED, NOT ACCORDINC- ARING MY- ILL NOT PE- OF THE DI- FOUND, AT THE SAME TI- ? THAT PREVAILING ANI- THE CONTINENT TO TH- OUS SPIRIT HAS TAINTED- I FORTIFY AS A MOUNT- IRE THE MOUNT- LCONSTRUCI- N, THE FRIEN- THEIR PROPS- NOUPLES TO V- UNCILS, HA, E, IN TRUTH- TO BE THE FAVORITE AV- E IMPROVEMENTS M- UCH ADMIRER, F- THIS SIDE, AS- LLY THE FRIEN- ONEFICES, I- OF JUSTICE AND TH- OUSLY WE MAY WISH- (HEY ARE IN 'SOME DEGR- 'VE BEEN ER- ILL NOT ALSO- 'EASING DIST- PUBLIC ENG- TLY, IF NOT P- NS, BY A FAC- TER EARLY M- NONE DESEF- 'ULAR GOV- TO THIS DAY- O HE IS ATT- 'AL DISEASE"- IS TO BE MORE ACCURATELY- VENTS NEVER FINDS HIMSEL- DUS VICE, HE WILL NOT FAIL- PROVIDES A PROPER CURE- 'AL DISEASE"- T FROM WE- AN CONST- INWARREL- CTED, COY- VATE FAITH, ONEFICES, I- NOR PARTI- AINTS HAC- FOUND, I- RGED ON- FOR MANY- ND ALARM- ECTS OF TH- DERSTAND- E CASE OF- H, HU EVERY- JUST REMEDY, I- 'RES, BUT IT CO- LD BE TO WISH TI- EXPEDIENT IS AS IT- EXERCISE IT, DIFFERE- AS AND H'S PASSION- 'H THEMSE- 'ES, THE DI- LE TO A U- 'ORMITY C- ENT AND V- 'UAL FACU- AND PROP- INFLUENCE- FERENT BY- S AND PARI- 'HT INTO D- 'DEGREES C- BRINGING RELIGION, CONCERNIN- RENT LEADERS AMBITIOUSLY CON- I INTERESTING TO THE HUMAN PASS- (DERED THEM- 'DISPOSED OPENSPITY C- ND FANCIF- MOST COM- CITIZENS? A- PROPOSED CL- 'JUSTICE OUGHT I- ARTY, OR, IN OTH- ID IN WHAT DEGRE- 2 AND THE MANUA- OF TAXES ON THE VAK- APS, NO I- 'SLATIVE A- JES OF JUST- 'E EVERY S- TO SAY TH- ALIGHTNE- 'LIC GOOD- 'HTENED ST- HOUT TAKI- O VIEW INL- TY MAY FIN- 'REGARDING- USES OF FACTION CANNOT BE RE- CONSISTS OF LESS THAN A MAJOR- S BY REGULAR VOTE, IT MAY CLOG TI- 'LENCE UN- 'THE POWER- OF THE CO- 'SACRIFICE- VATE RIGHT- THEN THE I- E THE DIFFERENT CLASSES O- PRIVATE DEBTS? IT IS A QUES- : BALANCE BETWEEN THEM- 'THE MOST POWERFUL FACT- SUPPLIES ON FOREIGN MARK- AND PRO- INS OF PRO- LATER OPP- 'HICH THE- WILL BE AB- 'ELM, NOR- NOT ALW- EMOTE COI- OF ANTI- D THAT REI- THE INFER- TO BE SOUGHT IN THE MEANS O- PUBLICAN PRINCIPLE, WHICH ENA- INVULSE THE SOCIETY; BUT IT WILL B- ITY IS INCL- 'ON, THE FA- 'OOD AND- E SUCH A F- OUR INDU- IE SAME TIA- LEFT ME ATIN- UJ CONCERNING IT, O THE CAUSES WHICH TH- IN ONE SIDE AND THE DESPO- ELVES THE JUDGES; AND THE M- 'OMESTIC MANUFACTURES BEEN- 'WOULD BEI- 'Y DECIDED B- THE APPOR- 'YES, THESE- TRAMPLE C- 'AKES, IT IS- 'IENT TO T- 'ENDER TH- UCH AN AI- HE IMMEDI- 'IT IS, THAT T- WHICH WE- 'ETS, IF A FAC- 'MAJORITY TO DEFEND ITS SINISTER- 'INABLE TO EXECUTE AND MASK- 'F POPULAR GOVERNMENT, 'IGHTS OF OTHER CITIZEN? RESERVE THE SPIRIT- 'E THE PUBL- OF POPULUS- 'THE DIFFERENT CLASSES O- PRIVATE DEBTS? IT IS A QUES- : BALANCE BETWEEN THEM- 'THE MOST POWERFUL FACT- SUPPLIES ON FOREIGN MARK- AND PRO- INS OF PRO- LATER OPP- 'HICH THE- WILL BE AB- 'ELM, NOR- NOT ALW- EMOTE COI- OF ANTI- D THAT REI- THE INFER- TO BE SOUGHT IN THE MEANS O- PUBLICAN PRINCIPLE, WHICH ENA- INVULSE THE SOCIETY; BUT IT WILL B- ITY IS INCL- 'ON, THE FA- 'OOD AND- E SUCH A F- OUR INDU- IE SAME TIA- LEFT ME ATIN-

A New Framework for Machine Learning and the Social Sciences

Justin Grimmer | Margaret E. Roberts | Brandon M. Stewart

DUSTIN S. STOLTZ / MARSHALL A. TAYLOR

Mapping Texts

COMPUTATIONAL TEXT ANALYSIS
FOR THE SOCIAL SCIENCES

O'REILLY®

Text Mining with R

A TIDY APPROACH

Julia Silge & David Robinson

HUMANITIES DATA ANALYSIS

Case Studies with Python

• • • FOLGERT KARSDORP • • •

• • • MIKE KESTEMONT • • •

• • • ALLEN RIDDELL • • •


```

root = lxml.etree.Element('sonnet')
# Add an author attribute to the root node
root.attrib['author'] = 'William Shakespeare'
# Add a year attribute to the root node
root.attrib['year'] = '1609'

for nb, line in enumerate(open('data/sonnets/116.txt')):
    line_node = lxml.etree.Element('line')
    # Add a line number attribute to each line node
    line_node.attrib['n'] = str(nb + 1)

    # Make different nodes for words and non-words
    word = ''
    for char in line.strip():
        if char.isalpha():
            word += char
        else:
            word_node = lxml.etree.Element('w')
            word_node.text = word
            line_node.append(word_node)
            word = ''

            char_node = lxml.etree.Element('c')
            char_node.text = char
            line_node.append(char_node)

    # don't forget last word:
    if word:
        word_node = lxml.etree.Element('w')
        word_node.text = word
        line_node.append(word_node)

    rhyme_node = lxml.etree.Element('rhyme')
    # We use xpath to find the rhyme node
    # and wrap it in a new node
    rhyme_node.append(line_node)
    line_node.replace(self, rhyme_node)

    root.append(rhyme_node)

# Add the volta
if nb == 8:

```

$$d(W_i, \mu_k) = \sum_{j=1}^J (W_{ij} - \mu_{jk})^2.$$

Using this measure of dissimilarity, we can assess the quality of any set of cluster centers μ . The k -means algorithm makes h each document is assigned to only one category (and thus only one element of the vector π_i equals 1 and the rest are 0). We can write the objective function—the function that reflects the quality of each solution and that our optimization algorithm will minimize—as:

$$f(\pi, \mu, W) = \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J \underbrace{(W_{ij} - \mu_{jk})^2}_{\text{dissimilarity measure}} \overbrace{\pi_{ik}}^{\text{Cluster indicator}}. \quad (12.1.2)$$

In words, the objective function measures the dissimilarity of documents from their assigned cluster centers, using the definition of dissimilarity from Equation 12.1.1.



To make it easy to use the functions `count_bigrams()` and `visualize_bigrams()` yourself, we've also reloaded the packages necessary for them.

```

library(dplyr)
library(tidyr)
library(tidytext)
library(ggplot2)
library(igraph)
library(ggraph)

count_bigrams <- function(dataset) {
  dataset %>%
    unnest_tokens(bigram, text, token = "ngrams", n = 2) %>%
    separate(bigram, c("word1", "word2"), sep = " ") %>%
    filter(!word1 %in% stop_words$word,
           !word2 %in% stop_words$word) %>%
    count(word1, word2, sort = TRUE)
}

visualize_bigrams <- function(bigrams) {
  set.seed(2016)
  a <- grid::arrow(type = "closed", length = unit(1, "cm"))

  bigrams %>%
    graph_from_data_frame() %>%
    ggraph(layout = "fr") +
    geom_edge_link(aes(edge_alpha = n), show.legend = FALSE) +
    geom_node_point(color = "lightblue", size = 5) +
    geom_node_text(aes(label = name), vjust = 1, hjust = 0.5) +
    theme_void()
}

```

```

x1_noise <- c(2, 4, 2, 1, 1)
x2_puppy <- c(0, 1, 4, 3, 2)
y_happy <- c(1, 3, 6, 7, 5)

## center our predictors
c1_noise <- x1_noise - mean(x1_noise)
c2_puppy <- x2_puppy - mean(x2_puppy)

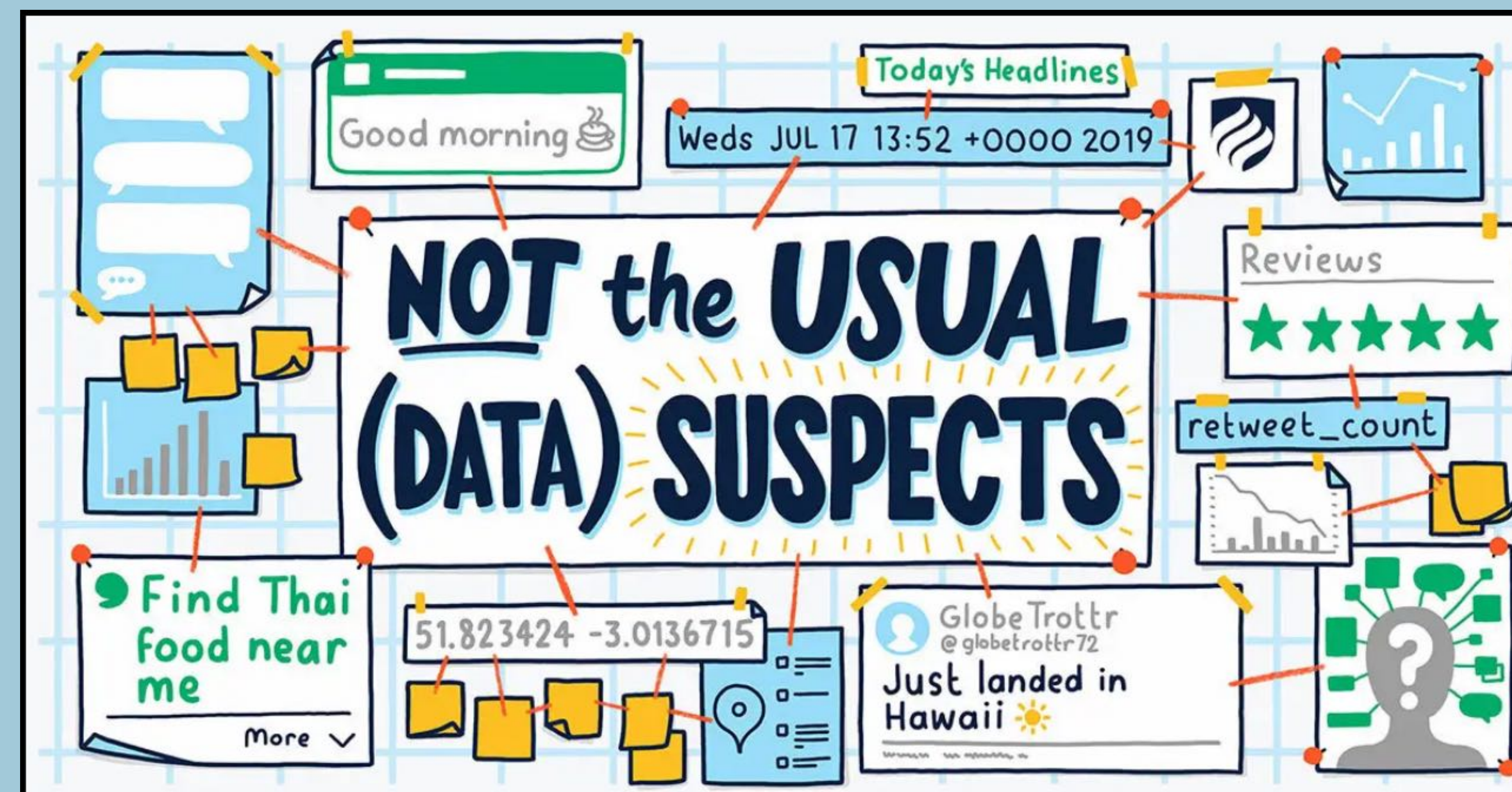
## dot products
dot_x1 <- sum(c1_noise^2)
dot_x2 <- sum(c2_puppy^2)
doty_x1 <- sum(y_happy * c1_noise)
doty_x2 <- sum(y_happy * c2_puppy)
dot_x1x2 <- sum(c1_noise * c2_puppy)

## beta 1
beta1 <- ((dot_x2 * doty_x1) - (dot_x1x2 * doty_x2)) /
          ((dot_x1) * (dot_x2) - (dot_x1x2)^2)

## beta 2
beta2 <- ((dot_x1 * doty_x2) - (dot_x1x2 * doty_x1)) /
          ((dot_x1) * (dot_x2) - (dot_x1x2)^2)

beta1
beta2

```

TEXT AS DATA

In the Humanities and Social Sciences

Consider that 90% of today's data has been created in just the last two years; Every day we create 2.5 quintillion bytes of data; Every 60 seconds, over 72 hours of video are uploaded to YouTube, over 200,000 Instagram posts are shared, and over 200 million emails are sent; And 90% of this data is unstructured.

How do we make sense of all this *noise* to help answer fundamental questions in the humanities and social sciences?

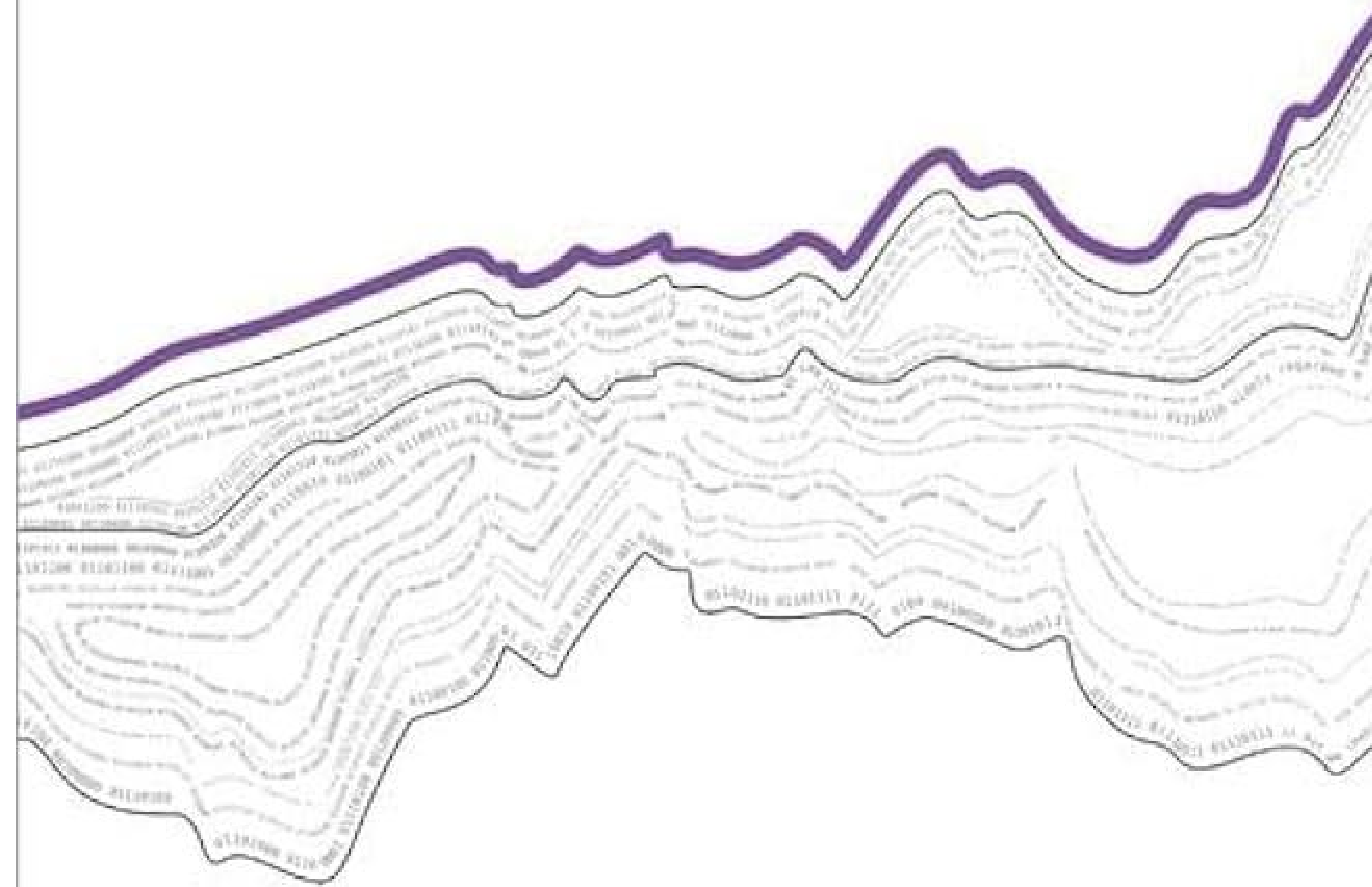
Join collaborative research class next fall exploring the potential of new technologies for analyzing texts – like books, essays, speeches, social media posts, lyrics, etc.

SOC 390R
FALL 2024

Interested?

Come to a brief meeting on Tuesday, March 26 at 1:00 in VZN 263
Or email Professor Roger Baumann (baumann@hope.edu)

DUSTIN S. STOLTZ / MARSHALL A. TAYLOR



Mapping Texts

COMPUTATIONAL TEXT ANALYSIS
FOR THE SOCIAL SCIENCES

