Racial Tropes in the Foreign Policy Bureaucracy: A Computational Text Analysis

Online Appendix

February 24, 2024

Contents

1 The President's Daily Brief **2**

1.1 Origins of the President's Daily Brief . . . 2

1.2 Production Process for the PDB . . . 3

1.3 PDB Processing . . . 4

2 Measuring Tropes **6**

2.1 Method 1: Trope Dictionaries . . . 6

2.2 Method 2: Supervised Learning . . . 11

2.2.1 Qualitative Coding Procedure . . . 11

2.2.2 Example Codings . . . 14

2.2.3 Predictive Process . . . 18

2.2.4 Model Metrics . . . 19

2.2.5 False Positives and Negatives . . . 26

3 Quantitative Descriptive Statistics **31**

4 Full Results and Robustness Checks **33**

4.1 Full Results . . . 33

4.2 Presidents and Time . . . 35

4.3 Entry Topics with a Structural Topic Model . . . 37

4.4 Including Leader Tenure . . . 45

4.5 Only Decolonized States . . . 47

4.6 More on Regional Variation . . . 49

4.7 More on Temporal Variation . . . 52

1 The President's Daily Brief

The President's Daily Brief (PDB) is a distillation of intelligence information, updates, and analysis that the intelligence community deems worthy of the president's attention. The PDB includes information obtained through all potential forms of intelligence (human, signals, open-source, and the like). As such, it is considered an "all-source" document.

1.1 Origins of the President's Daily Brief

Much of the information regarding the PDB's origins that we report below come from two sources. One is a book by David Priess, which we cite throughout our main text (Priess, 2016). The second is a document published by the CIA when the organization released its first tranche of PDBs in 2015 (Mansfield, 2015).

The intelligence bureaucracy produces hundreds of reports daily – far too much for a single individual to read, and from a wide array of sources that do not necessarily coordinate with one another. Following World War II (and particularly with the attack on Pearl Harbor in mind), President Harry Truman initiated an effort to coordinate and consolidate information from the range of intelligence agencies in the federal government. This led to the establishment of the Central Intelligence Group (CIG) in January 1946. The CIG attempted to fulfill Truman's requests for intelligence coordination by producing the Daily Summary, which began in February 1946. The Summary continued after the CIG was replaced by the Central Intelligence Agency (CIA) in accordance with the National Security Act of 1947. This progenitor of the President's Daily Brief predominantly featured information from State Department sources (in 1949, a committee analyzing the intelligence system found that about 90% of these daily reports were based on State Department sources), did not feature much in terms of signals intelligence because the CIA did not provide much material, and only provided statements of fact without analysis.

In February 1951, the Summary was retired and supplanted by the Current Intelligence Bulletin. The Bulletin was an important development and advancement; it was the first all-source (based on all possible sources of intelligence) document and featured analytic commentary. During the Eisenhower administration, the Bulletin's relevance as a daily supply of intelligence to the president faded. Eisenhower preferred to learn and discuss developments during his regularly-scheduled National Security Council (NSC) meetings. Until the end of his presidency, Eisenhower preferred receiving a constant and voluminous stream of reports and briefing materials that would be the basis of discussion in NSC gatherings. The CIA worked to satisfy these demands. Simultaneously, however, the CIA attempted to regain Eisenhower's interest in the Current Intelligence Bulletin by overhauling it. The resulting *Central* Intelligence Bulletin deepened the level of information, analysis, and graphics provided to the reader. The new Bulletin failed to interest Eisenhower but did receive wider circulation. This wider circulation, however, also prevented the Bulletin from containing the most sensitive intelligence.

When Kennedy entered office, his relative lack of experience and interest in foreign policy made him unwilling to read the lengthy bureaucratic materials that Eisenhower had sought and that the CIA had produced in response. (Indeed, at the start of his presidency, Kennedy obtained most of his news on foreign developments by reading newspapers, which were written in a more concise and casual manner that aligned with his preferences.) The disastrous Bay of Pigs invasion in 1961, as well as Kennedy's apparent inability to pay attention during extended in-person intelligence briefings, forced the president and his advisers to reconsider their approach to intelligence reporting. Chester Clifton, who served as a senior military aide and Kennedy's daily briefer, was not an intelligence analyst and struggled to filter each day's large stack of reports into a more focused summation of key items. At Clifton's direction, Huntington Sheldon – then Director of the Office of Current Intelligence in the CIA – recruited analyst Richard Lehman to produce a new document that would contain information from the most classified of sources but also be short and simply written. Their efforts generated the President's Intelligence Checklist (PICL). Kennedy expressed great satisfaction when receiving the first PICL on June 17, 1961.

Following Kennedy's assassination, Johnson showed disinterest in the PICL, potentially seeing it as a legacy of the previous administration. The PICL had also gradually grown in length and complexity. As a result, Johnson – who appeared not to be as interested in reading longer documents – relied more heavily on the CIA's twice-weekly summary of global events. This document, the President's Intelligence Review, looked much more like early PICLs but could not offer as much information due to its twice-weekly nature. CIA officials attempted to indirectly get information in the PICL to the president by expanding its distribution to more White House officials. However, the widening circulation also came at the expense of diluting the degree of sensitive information published in the document. Intelligence officials tackled these issues by producing a PICL replacement. This new all-source document appeared more formal, brought back highly classified intelligence, had extremely limited distribution, and was designed to make Johnson feel like it had been created specifically for him. This was the PDB, which was first published on December 1, 1964. Johnson quickly grew to appreciate the PDB.

1.2 Production Process for the PDB

The production process for the PDB has changed over time. During the time of the PICL, a couple analysts from a small circle of men would review the array of intelligence reports they were sent over the entire day, creating an initial draft the evening before the document was needed. Around 4 or 5 AM, an analyst would return and make any necessary updates, even adding new intelligence up until only an hour before the PDB would be delivered to the president (Kovar, 2000).

The PDB now undergoes a more systematic process involving more actors. The most recent comprehensive review of this process we could find comes from an article in 2008 (Johnson, 2008). As such, some specific details such as times may vary across years and administrations from what the 2008 piece states, but the broader process should remain largely the same.

The cycle (at least in the early 2000s) begins at 8:30 AM the day before a PDB is actually distributed. CIA officials orchestrate meetings in which they determine what subjects will likely be discussed in the PDB. These decisions are made based on pressing events, requests senior policymakers or the president may have previously made for additional information, as well as assessments of what issues are likely to show up in leading media publications. As topics are determined, analysts across the entirety of the intelligence bureaucracy (see Table A1 below) are contacted to contribute relevant information on the narrowing list of issues. Analysts and offices submit this requested information between 10:45 AM and the early evening. Throughout this stage, potential snippets of information are circulated among offices to collect revisions, corrections, and amendments. A drafting group compiles, shortens, and synthesizes this information into an initial draft that is reviewed by the CIA's Deputy Director for Intelligence and the Director of the CIA around 8:00 PM. Their revisions are incorporated into the PDB draft. The PDB continues to be updated with the latest information and is readied for publication overnight. The final PDB is finalized and printed (or uploaded to a tablet) at 5:30 AM and is carefully distributed to relevant recipients between 6 and 9 AM.

The PDB was originally produced under the auspices of the Central Intelligence Agency. It is now administered by the Office of Director of National Intelligence, which was established in 2005 (following the terrorist attacks of September 11, 2001) to further coordinate all intelligence agencies. As of 2023, the United States intelligence community is comprised of 18 organizations, which are listed in Table A1 in chronological order. Each organization features subsidiary intelligence groups and offices that provide information to the appropriate director, who in turn makes decisions about what material to submit for potential inclusion in the PDB.

Table A1: List of organizations in the US intelligence community.

Organization	**Year Est.**	**Department**
Office of Naval Intelligence (ONI)	1882	Defense
Coast Guard Intelligence (CGI)	1915	Defense
Bureau of Intelligence and Research (INR)	1945	State
Air Force Intelligence, Surveillance, and Reconnaissance Agency (ISR)	1948	Defense
Central Intelligence Agency (CIA)	1947	Independent
National Security Agency (NSA)	1952	Defense
Defense Intelligence Agency (DIA)	1961	Defense
National Reconnaissance Office (NRO)	1961	Defense
Army Intelligence and Security Command (INSCOM)	1977	Defense
Office of Intelligence and Counter-intelligence (OICI)	1977	Energy
Marine Corps Intelligence (MCI)	1978	Defense
National Geospatial-Intelligence Agency (NGA)	1996	Defense
Office of Intelligence and Analysis (OIA)	2004	Treasury
FBI Intelligence Branch (IB)	2005	Justice
Office of the Director of National Intelligence (ODNI)	2005	Independent
Office of National Security Intelligence (ONSI)	2006	Justice
Office of Intelligence and Analysis (I&A)	2007	Homeland Security
National Space Intelligence Center (NSIC)	2020	Defense

1.3 PDB Processing

The first seventeen years of the President's Daily Brief (PDB) were declassified by the Central Intelligence Agency in 2016. A team of research assistants downloaded these briefs, which were posted online in PDF form, and then used optical character recognition (OCR) software to convert the PDFs into machine-readable text. Throughout this process, research assistants were instructed to delineate sections of the PDB by entry – a natural unit of analysis in the document across its entire existence. They were also asked to split each entry by title (if applicable) and the main text.

Figure A1 illustrates how a page of the PDB from February 8, 1966 is split at the entry level. We identify each discrete entry in a PDB, further dividing entries by their title (if applicable) and text. The country or countries mentioned in an entry are identified using information from both the title and text.

Figure A1: Page of the PDB from February 8, 1966, with entry segmentation shown.

DAILY BRIEF
8 FEBRUARY 1966

1. North Vietnam	The list of North Vietnamese diplomats who have returned to Hanoi is lengthening. Those known to have returned are Hanoi's representatives to France, Poland, Ghana, Guinea, Tanzania, and India. In addition, the ambassadors to Moscow, Peking, and Havana may also be home. Hanoi is probably conducting a general foreign policy review. Such reviews are known to have taken place at about this time of year in 1962 and 1964. These earlier meetings were not followed by any major alterations in foreign policy.
2. Uganda	The Ugandan parliament has forced the suspension of Prime Minister Obote's leading supporter, Colonel Idi Amin, from his post as army chief of staff. Amin is under investigation for involvement in the embezzlement of Congolese rebel gold. This is a severe setback for the radical Obote and drastically reduces his chances for re-election later this year. The mounting conflict between moderates and leftists sparked reports yesterday that a coup was imminent. The embassy reported this morning that the Kampala area was calm, but that political leaders were operating on an "anything can happen" basis.

The output of the OCR software contains errors, particularly given the imperfect nature of the PDFs and the fact that they are scans of decades-old documents. The PDBs also feature redactions, which mean that specific words and phrases that could identify still-sensitive information were concealed in the PDFs that the CIA released. As such, research assistants were also tasked with manually reviewing the OCR output and making any necessary corrections.

We create a dictionary of terms to identify mentions of countries in the PDB. For each country, we collect historical names, names of capitals, names of highly-populated cities, standard abbreviations of cities (such as UAE for United Arab Emirates), and demonyms ("Indonesian" for Indonesia). This dictionary is then applied to each PDB entry to identify mentions of any countries. For each country mentioned, we produce an additional entry in our data. An entry that mentions three different countries is therefore translated into three rows in our data, where all three rows are identical in text and other metadata except for a variable indicating a country.

To identify whether leaders are mentioned, we use the Archigos dataset. For each entry, we use Archigos to determine the name of the leader in power for a country on the specific date of the entry. If the leader's name shows up in the entry, this is coded as an instance of a leader being mentioned. The robustness check in Appendix 4.4 uses (logged) leader tenure in days. To calculate this, we identify the leader in the country and the difference between the day of the PDB entry and the day they initially took power.

2 Measuring Tropes

We create quantitative measures of tropes in the PDB using two methods. One involves the creation of a dictionary of synonyms, while the other involves a series of supervised machine learning models.

2.1 Method 1: Trope Dictionaries

A large part of our analyses relies upon a dictionary-based technique to identify instances of racial tropes. This dictionary was constructed by first identifying sets of words we believed were related to each trope; see Table A2. We then used the open-source Moby thesaurus to find all words and phrases synonymous with these initial terms. This process produced lists of words plausibly associated with each trope. From these lists, we removed all terms that occurred extremely frequently (at the 95th percentile or higher), as these often reflected common terms that were too broad to usefully identify our concepts of interest. For example, some of the most commonly used synonyms for irrationality were "change" and "possible." For animal analogies, we see similar issues with terms like "attack" and "enemy."

Table A2: Initial terms used to construct dictionaries of synonyms.

Trope	**Initial words**
Infantilization	childish, baby, naive, whining, callow
Animal analogies	animal, animalistic, savage, [animal names], [animal actions]
Belligerence	belligerent, hostile, angry
Irrationality	chaotic, confused, crazy, emotional, insane, irate, irrational, unstable, volatile

This process ultimately produces the dictionaries we use to track the invocation of potential tropes in our text data. These dictionaries are supplied in Tables A3 (infantilization), A4 (animal analogies), A5 (belligerence), and A6 (irrationality).

Table A3: Dictionary of terms for infantilization.

acut	debut	grievanc	juvenil	plain	sketchi	unfamiliar
adolesc	deceiv	grope	keen	plaintiv	skirt	unfamiliar with
awkward	decrepit	grous	kick	pocket	slip	unguard
babe	defect	grumbl	lamb	precious	small-scal	uninform
babi	defici	half-bak	lament	preschool	snipe	unreserv
backward	delud	handi	lightweight	probation	sob	unseason
beef	diminut	hon	love	puppet	soft	unsophist
blatant	dissent	honey	maiden	querul	sop	unstudi
bluff	doll	hoodwink	mini	raw	sophomor	unsur
blunt	duck	howl	miniatur	raw recruit	sorrow	untouch
bud	easi	humor	minuscul	retard	spoil	untri
cater to	embryon	immatur	miss	ripen	squeak	unus
chicken	empti	in arm	mourn	scant	suscept	unus to
child	entrant	in arrear	mous	scanti	sweet	unwari
childish	firsthand	in short suppli	naiv	schoolgirl	teenag	vest-pocket
coddl	flame	inadequ	new to	scold	tender	virgin
compact	fledgl	inan	newcom	screech	tomato	whistl
complaint	foolish	incomplet	nonent	senil	tot	youngest
cotton	frail	indulg	novic	shi	toy	youngster
coward	frank	inexperienc	outspoken	shrill	traine	
creaki	fret	infant	peevish	sigh	trust	
cri	girl	innoc	pet	silli	unaffect	
dame	gratifi	intact	petul	simpl	underdevelop	
dear	green	juici	piec	sincer	undevelop	
deb	greenhorn	junior	pipe	single-mind	unenlighten	

Table A4: Dictionary of terms for animal analogies.

abus	chatter	fierc	hopper	molest	rat	swim
afflict	chauvinist	fire-eat	hornet	monkey	rave	swine
aggriev	chicken	firebrand	hors	monster	ray	swoop
agon	child	flamboy	hothead	mortal	relentless	taint
ambl	chip	flap	hound	mous	rip	tear
anim	citat	fledgl	howl	mutton	roar	termag
animal	clan	flesh	impair	nest	rook	tiger
ant	claw	flock	implac	nibbl	roost	tod
antagonist	cluster	flyer	impolit	nurseri	rough up	tongu
ape	clutch	fowl	incendiari	obstin	rude	torment
arc	coars	fox	infant	ostrich	rug	tortur
ass	cock	fractur	infect	otter	ruin	traumat
ave	cod	fray	infuri	ounc	ruptur	trigger-happi
bacon	congreg	fret	inhuman	outrag	ruthless	trot
badger	convuls	frog	inim	ox	saber-rattl	trucul
bale	coron	fume	injur	parrot	sack	tyranni
barbar	cow	furi	insect	pernici	sanguin	unbroken
bask	crane	furious	instinct	persecut	savag	uncontrol
bat	crawl	fuzzi	internecin	pet	scath	uncultiv
batter	crawler	gall	jingoist	pig	scold	unfriend
beast	creatur	gallop	juvenil	pigeon	scotch	ungentl
beef	crocodil	gam	kaleidoscop	pillag	scrape	unkind
bellicos	crow	gaze	karakul	piti	scratch	unrel
belliger	cruel	giant	kettl	play havoc with	scurri	vandal
bird	cur	glean	killer	play hob with	seal	veal
bite	curs	goat	kindl	plump	sheep	vermin
black sheep	dazzl	goon	knock about	pod	shiver	vicious
bloat	deceit	grim	ladi	poison	shoal	virul
bloodi	demon	grind	lamb	pollut	shrewd	warlik
bloodthirsti	deprav	grist	lament	pork	singular	weasel
bloom	devil	gross	laps	poultri	sire	wedg
bouquet	diabol	grous	lay wast	pounc	skin	whoop
bowl	dirti	grumbl	leash	prejudic	skipper	wiggl
brood	disadvantag	gull	lem	primit	skunk	wild
bruis	distress	gulp	lethal	prowl	slash	wild beast
brutal	dog	ham	lion	pugnaci	slaughter	wildcat
buck	dole	hardnos	lobster	punctur	slit	witch
buckl	doom	hare	locust	pup	smack	wolf
buffalo	dove	harrow	loot	rabbit	snail	worm
buffet	Draconian	hawk	maim	rabbl	sneak	wrack
bull	dragon	hawkish	mammoth	rabid	soar	wreck
busy	drone	herd	manhandl	rack	sow	wrench
buzz	duck	hind	martyr	raft	spat	wriggl
camel	dule	hive	maul	rag	squab	wring
cannib	enrag	hob	menac	rage	stab	yoke
caravan	exalt	hog	merciless	ram	stag	
cat	fatal	hood	militarist	ramp	steer	
cattl	feroci	hoodlum	mischief	rampag	streak	
cauldron	ferret	hop	mishandl	rant	stud	
chafe	fever	hop mad	mistreat	rape	swarm	

Table A5: Dictionary of terms for belligerence.

acid	caustic	dirti	grate	malici	rile	troublesom
acrimoni	chafe	disaffect	harm	martial	rioter	trucul
advers	chaotic	disagre	hatchet man	mif	rival	tumultu
adversari	chauvinist	disapprov	hate	milit	roil	turbul
aggrav	cloudi	discont	hawk	militarist	rough	uncoop
aggress	cold	discord	hawkish	miser	rowdi	unfavor
alien	collid	discrep	heat	negat	ruffl	unfriend
anarch	combat	disench	hood	nettl	saber-rattl	unpropiti
anger	compens	disharmoni	hoodlum	not easi	sanguin	unsympathet
angri	competit	disinclin	hooligan	obstin	savag	untoward
annoy	competitor	displeas	hostil	oppon	sensit	up in arm
antagonist	con	disproportion	hot	out of temper	set against	upset
antipathet	conflict	disputati	ill	partisan	sinist	variant
ardent	confront	dissid	in opposit	pervers	smart	venom
argument	contend	disturb	incens	piqu	soldier	vex
at loggerhead	contenti	diverg	incompat	polar	sore	virul
at odd	contest	divis	inconsist	polem	sour	vitriol
at varianc	contradict	enforc	indign	provok	spite	warlik
at war	contradictori	enrag	inflam	pugnaci	squar off	wild
avers	contradistinct	exasper	infuri	put off	storm	wild-ey
balanc	contrari	eyebal to eyebal	inhospit	put out	stormi	work up
battl	contrast	feroci	inim	quarrel	stress	wrangl
bellicos	controversi	fester	invad	rage	strong-arm	wrath
belliger	convers	fierc	irasc	raini	strong arm	wretch
bicker	corros	fieri	irat	rancor	struggler	wrought-up
bitter	counter	fight cock	ire	rankl	swashbuckl	wrought up
bloodi	counteract	fighter	irk	rave	sword	
bloodthirsti	counterbalanc	foul	irrit	raw	swordplay	
bluster	counterpois	fractious	jangl	reactionari	temper	
blusteri	countervail	frantic	jar	recalcitr	tempestu	
bother	cross	frenzi	jingoist	red	tender	
breakaway	cyclon	fume	livid	resent	thug	
bug	dead against	furious	loath	revers	tick off	
bulli	detriment	gall	mad	revolutionari	tough	
burn	dim	goon	madden	rigor	trigger-happi	

Table A6: Dictionary of terms for irrationality.

abnorm	crunch	fli	indefinit	negat	shaken	undefin
absurd	cuckoo	flicker	indetermin	nervous	shaki	undepend
adrift	damn	fluctuat	indign	noisi	shallow	undetermin
afflict	dark	fluid	indiscreet	nonperman	shi	undisciplin
afloat	decim	fluster	indistinct	nonpluss	shook	uneasi
agit	decrepit	foggi	indistinguish	nonsens	short-liv	unequ
airi	delud	fond	inept	nonspecif	shot	uneven
ajar	dement	fool	infatu	not bright	shot through	unforese
altern	demur	foolish	infinit	not follow	shuffl	unhealthi
ambigu	derang	fraction	infirm	not right	sick	unorgan
amorph	desultori	fragil	inflamm	numer	silli	unorthodox
anarch	deviat	frail	infuri	nut	simpl	unpredict
anarchi	devious	frantic	injudici	obscur	skittish	unrealist
anarchist	die	frenet	innat	odd	slipperi	unreason
anger	differenti	frenzi	inordin	of soul	snarl	unreli
angri	digit	fugit	insan	off the track	sophist	unrestrain
ape	dim	furious	insecur	one-sid	sore	unruli
ardent	discomfit	fuss	integr	ordin	soul	unsaf
astray	discomfort	futil	intellectu	out of it	spineless	unsettl
asymmetr	disconcert	fuzzi	intric	outrag	spiritu	unsound
at a loss	disintegr	grate	invalid	overeag	sporad	unspecifi
at fever pitch	dismay	guess	irasc	overenthusiast	sprung	unstabl
at sea	disord	gull	irat	overzeal	stir	unsteadi
avid	disorgan	gut	ire	pair	storm	unstuck
babbl	disori	haggard	irrat	pale	stormi	unsubstanti
backward	disposit	hair-trigg	irresolut	passion	strang	unsur
baffl	disquiet	half-bak	irrespons	peril	stupid	unthink
banana	distort	haphazard	irrit	perish	subnorm	untrustworthi
bat	distract	harsh	jangl	perplex	subtl	unwis
befuddl	distress	hazard	jar	pluralist	superfici	unwork
bemus	dither	hazi	jostl	pointless	suscept	upset
bent	diverg	heat	jumbl	possess	sweep	vacil
beset	diversifi	hectic	kaleidoscop	precari	sympathet	vagu
bewild	downi	heel	keen	prick	tangl	vari
big	dread	high emot	keen on	protean	temperament	variabl
bizarr	dubious	hothead	knot	provision	tempestu	veil
blur	dull	hung up	labyrinthin	put off	tempor	versatil
blurri	eager	hyster	lax	put out	temporari	viscer
bluster	edgi	ignor	lean	puzzl	tender	volatil
blusteri	elabor	ill-advis	livid	queer	tens	wander
bother	embarrass	ill-consid	loos	rabid	tentat	wanton
bow	emot	ill-defin	lopsid	rag	thoughtless	warm
brittl	emotion	imaginari	ludicr	rage	tick off	waver
bubbl	entangl	impair	lunat	rambl	ticklish	wayward
bug	enthusiast	impass	mad	random	timid	weak
buoyant	ephemer	impetu	madden	rant	topsy-turvi	weightless
Byzantin	eros	implic	maniac	rash	tortuous	weird
cant	errat	impolit	manic	ration	touch	wet
caprici	erupt	imponder	mat	rattl	touchi	wild
cardin	evapor	imposs	maudlin	rave	toy	wild-ey
certifi	excit	impract	maze	reciproc	transient	wishy-washi
chagrin	expans	imprecis	meander	reckless	transit	without reason
chanci	explos	imprud	medley	resili	transitori	wobbl
chaotic	extravag	impuls	mental	restless	treacher	work up
characterist	fade	inaccur	mercuri	retard	tricki	wrath
clamor	faint	inan	messi	ridicul	tumultu	wrought-up
color	fallaci	incens	miscellan	riski	turbul	yeasti
confound	fanat	inchoat	misguid	rotten	turn around	zealot
confus	fantast	incoher	mislead	rough	twist	zealous
conscious	faulti	inconclus	mix up	roundabout	unaccount	
contradictori	feebl	inconsequenti	mobil	rove	unbalanc	
corrupt	fervent	inconsist	momentari	ruffl	uncertain	
coy	feverish	inconspicu	monstrous	scatter	unclear	
crack	fit	inconst	mortal	senseless	uncomfort	
crazi	flaw	incred	motley	sentiment	uncontrol	
crumbl	fleet	indecis	muddl	shadowi	undecid	

2.2 Method 2: Supervised Learning

2.2.1 Qualitative Coding Procedure

To permit qualitative coding of PDB entries for tropes, we developed a comprehensive codebook as a basic set of instructions for how to code each PDB entry in the data set. The codebook indicated our tropes of interest, provided some terms or phrases that could plausibly indicate them, and supplied a series of examples that we as co-authors had previously coded together. This codebook was provided to a series of research assistants (RAs), as well as a randomly sampled subset of our PDB data comprising 2,340 unique entries.

We did not tell the coders the research questions or broader theory motivating the paper. Instead, we instructed them that their task was to explore PDB entries to determine the presence of one or more of what we called "themes" in the codebook. Table A7 replicate the core instructions we provided to the RAs for each theme. For each entry, the RA would qualitatively review the text and provide a binary coding of whether a specific theme was present in the text (1) or not (0). We also noted the possibility that certain entries would satisfy more than one of the themes, so one entry might be coded as having multiple themes present. We also noted that it would be likely that many entries were coded as 0 for most if not all of the possible themes. We stressed to the RAs that they were not simply on the hunt for individual words or only the words mentioned in our examples. For instance, we noted that even entries that suggest a lack of rationality without saying "irrationality," for example, should be coded as 1 for irrationality. RAs were invited to ask the graduate RA and the authors about any entries they were unsure about.

Qualitative coding took place over two separate phases. The first round of coding, which took place between January and April 2021, was the most labor-intensive, involving three undergraduate RAs and one graduate RA. The vast majority of coding took place in this initial phase. During an initial orientation phase, team members were presented with a sample of 102 randomly selected entries and coded these individually. Intercoder reliability was relatively high, with the initial average pairwise Pearson correlations between all four coders being 0.81 (infantilization), 0.70 (animal analogy), 0.73 (belligerence), and 0.74 (irrationality). After discussing and resolving discrepant codings on this initial sample of 102, the team members individually coded additional entries from the larger sample. After the orientation phase, only one individual was responsible for coding the entry.

A second and more abbreviated round of coding took place from December 2022 to January 2023. An additional graduate RA was recruited to (1) code some additional entries for irrationality (where our first-round data had relatively few positive cases), and (2) to review some previous codings of irrationality and belligerence.

Table A8 lists the total number of positive examples that RAs identified for each trope.

Table A7: Instructions provided to coders for each theme.

Trope	**Instructions**
Infantilization	This is language that refers to individuals, groups of people, entire countries, or other actors as children or in a way that denies actors' maturity in age or experience. Often these terms refer to the inability to stay calm or collected, implying foreign leaders had "tantrums" or similar behavior we associate with toddlers and young children. If you see infantilizing language, code the INF variable as 1; otherwise code as 0.
Animal analogies	These passages include language that either compares actors to animals or uses idiomatic language that centers animals, refers to conditions or movements we associate with animals (such as rabies), or refers to wildlife/nature more generally. These passages are not mentioning animals in reality. For those passages that use metaphors or idiomatic language centering animals, code this passage as a 1; otherwise 0.
Belligerence	Another common emotional attribute is belligerence. Belligerence will often be tied to violence and these entries will discuss emotions having to do with anger, frustration, bitterness, or belligerence. Within this category there are two interesting metaphorical patterns: metaphors of heat/explosions and weather metaphors. These entries do not have to include metaphors to be coded as attributing a belligerent attitude to actors. Any time entries imply belligerence/anger/generally violent emotions, including those entries that utilize the metaphor examples below, code the BELLIG variable as 1; otherwise 0.
Irrationality	Next, we are interested in entries that discuss irrationality (see rationality below). If an entry is attributing irrationality, PDBs will discuss an actor as irrational or crazy directly, implying actors suffer from mental illness or paranoia. PDB entries might also imply irrationality by claiming not to understand why an actor would take a certain action and implying that understanding certain actors' motives is impossible. In other words, irrationality implies a flawed decision-making process on the part of an actor. If rationalism assumes actors take into account all possible actions, the possible outcomes, and the probabilities of success associated with each action, actors should choose the most beneficial action. In the case of irrationality, entries assume that leaders have either not followed this process, have erred in this process, or are incapable of making decisions rationally. When attribution of irrationality is present, code the IRR variable for the entry as 1; otherwise 0.

Table A8: Total number of positive hand-coded examples for each trope.

Trope	**Positive examples**
Infantilization	137
Animal analogy	99
Belligerence	373
Irrationality	97

2.2.2 Example Codings

Table A9 provides examples of PDB entries that address the same country and are coded positive and negative for each of the trope themes we assess.

Table A9: Positive and negative examples of trope presence in PDB entries for the same country.

Trope	Positive Example	Negative Example
Infantilization	Laos: Ambassador Unger thought both Kouprasith and Siho were acting like 'badly frightened little boys.' (DOC_0005959123, 20 April 1964)	Laos: Souvanna says that the Pathet Lao by their offensive in central Laos which has resulted in the capture of Na Kay, have wrecked his efforts to pull the three government factions together. He has suspended preparation for further meetings in the Plains des Jarres. (DOC_0005996809, 1 February 1964)
Animal analogy	Congo: Gizenga has reversed himself and now says he will go to Leopoldville to face parliamentary charges. This came after General Lundula called his hand and threatened to put him under arrest and after advice from the Egyptians and Yugoslavs to give up. He may still try to wriggle away - he seems to have a genuine fear of meeting Lumumba's fate if he goes to Leopoldville - but whatever he does, his political position is now badly crippled. (DOC_0005992137, 12 January 1962)	Congo: B. In Stanleyville, Gizenga's position seems to be waning somewhat. Cairo has again urged him to return to Leopoldville contending that, otherwise, he will be read out of the central government soon. (DOC_0005992119, 2 January 1962)
Belligerence	Dominican Republic: Gen. Sanchez, the nominal head of the air force was, however, in a bellicose and intransigent mood yesterday and has not great regard for Balaguer. (DOC_0005992043, 19 November 1961)	Dominican Republic: Our Consul reports that after the bad fright they suffered in last week's crackdown the opposition is beginning to regain its composure. On the government side both Ramfis and Balaguer are having trouble with old-guard members of the Trujillo family and the armed forces and appealed for token US support. (DOC_0005958993, 9 August 1961)
Irrationality	Congo: Mobutu's pursuit of grandeur is carrying him to new follies that make some of his other Alice-in-Wonderland performances seem almost rational by comparison. (DOC_0005976305, 15 August 1968)	Congo (Brazzaville): Chances of early military intervention by Kinshasa have faded. Mobutu is junketing around various North African capitals and seems to have lost his opportunity. Brazzaville's new leader, Ngouabi, appears to be consolidating his power position handily. (DOC_0005976369, 21 September 1968)

Below, we provide four additional positive examples that research assistants identified for each trope.

Infantilization

- 13. British Guiana: We expect a close vote on 7 December and Jagan's party is likely to win a plurality but not a majority under the new proportional representation system. The prospective anti-Jagan coalition partners are already squabbling. (DOC_0005967390, 27 November 1964)

- 5. Colombia: President Valencia's National Front government has come under heavy fire. The coalition is beset with factionalism and bickering, the antiguerrilla campaign has bogged down, and a host of economic problems remains. Political leaders on all sides are now saying that Valencia must go. In a situation like this a military takeover is always possibility. (DOC_0005959454, 29 September 1964)

- 8. Nigeria: Verbal jousting between Supreme Commander Gowon and Eastern governor Ojukwu is threatening to spill over into military action. Ojukwu seems to be gearing his people up for an early attack by federal forces. He is also said to be dickering with some tribal leaders from the Western region who are getting fed up with the presence of Gowon's mainly northern army. (DOC_0005968826, 9 March 1967)

- Arab States: The Arab foreign ministers' conference in Kuwait, which ended Saturday, did nothing to alleviate the disarray among Arab states. The participants were not able to agree even on a joint communique, much less make any progress on such issues as Jordan's long-standing quarrel with the fedayeen. The Jordanians refused to go along with the terrorists' demand that they be allowed to operate everywhere, including Jordan. Efforts by some ministers to promote an Arab summit meeting also apparently got nowhere. (DOC_0005993635, 20 November 1972)

Animal analogies

- 1. Indian-Chinese Communist Border Dispute: Nehru's disclosure of more Chinese incursions during the past 18 months will increase pressure on his government to strengthen military forces on the border. With national elections due next February, the Indians may put on a show of strength at the frontier even though this could well lead to further clashes with Chinese border units. Soviet officials continue to view Chinese bull-headedness on this issue as a grave error which is driving Nehru into the arms of the West. (DOC_0005992046, 21 November 1961)

- 1. South Vietnam: Ky, whose desire to run for president is becoming ever more compelling is worrying now over how best to weasel out of his earlier pledge to back Chief of State Thieu for the office. unless Thieu soon decides on his own not to make the race the two men will be obliged to have it out between them. (DOC_0005968734, 16 January 1967)

- Europe: Mayor Schuetz leaves Berlin today on a visit to Poland which has all the overtones of a typical cat-and-mouse game between European statesmen – and between West German domestic political rivals as well. The various participants in the game have a wide assortment of aims. (DOC_0005976827, 14 June 1969)

- USSR-China: Brezhnev lent his personal prestige to the current campaign against China yesterday by condemning "the practice of Maoism." In a speech in Alma Ata, not far from the Chinese border the Soviet leader gave a gloomy assessment of relations with Peking and echoed the strident themes of recent Soviet propaganda. According to a TASS summary of the speech, Brezhnev placed special stress on Moscow's vague suggestion for an Asian collective security system, implying that China's "rabid anti-Sovietism" is the major obstacle to stability in Asia. Brezhnev himself first floated the Asian collective security proposal in a clearly anti-Chinese context four years ago. Since then Soviet pronouncements have plugged the idea periodically particularly since the signing of the Vietnam peace agreement early this year. (DOC_0005993904, 16 August 1973)

Belligerence

- C. In an interview with the Iranian Prime Minister last Friday, the Soviet Ambassador indirectly threatened strong Soviet action unless Iran withdrew from CENTO. Other Soviet diplomats in Iran have reportedly stated that the USSR will invoke the 1921 Treaty of Friendshipâ, which Moscow claims gives it the right to occupy Iran if Soviet security is threatened, in order to forestall CENTO aggression. (DOC_0005959036, 12 September 1961)

- 3. Nigeria: Security officials in Lagos have picked up scores of machetes from eastern tribal activists in the area. This lends some substance to persistent reports that these easterners were arming themselves for attacks against selected northerners. Any such violence would reinforce already strong northern secessionist sentiment. (DOC_0005968493, 26 August 1966)

- 10. Sierra Leone: The chief of the army has moved in to prevent the new prime minister from taking office. This brought on angry demonstrations yesterday and more disr turbances are likely today. In sum, a considerable period of instability seems to be in the cards for this West African member of the British Commonwealth. (DOC_0005973690, 22 March 1967)

- Egyptian workers and students demonstrating violently in Cairo yesterday over economic grievances called for a return to Nasirist socialism. Prime Miniter Hijazi was the chief object of the protesters' wrath, but President Sadat was also criticized for failing to bring about promised economic recovery after the October 1973 war. Discontent over shortages and inflation has been on the rise since last summer and this dissatisfaction has been compounded by restlessness over the pace of progress in peace negotiations. Additional disturbances over a wide range of issues are possible and, although President Sadat now intends to move against leftist agitators, yesterday's demonstration could give these elements and student malcontents further impetus. (DOC_0006007907, 2 January 1975)

Irrationality

- 4. Dominican opposition's approach to US Consulate: Two sources have reported to the Consulate that President Balaguer was unable to control the Trujillo family. Stating that immediate US armed intervention was the only solution, they proposed to begin a campaign of sabotage against US holdings in order to "force" intervention. The Consulate told them such action was stupid and tends to discount it as a possibility. (DOC_0005958912, 19 June 1961)

- D. Ghana: Yesterday's attempt on Nkrumah's life is virtually certain to bring forth another round of anti-US hysteria, as was the case in 1962. It will also add impetus to a purge and government reorganization which pro-Communists have been urging on Nkrumah. (DOC_0005996752, 3 January 1964)

- 4. South Korea: Korea's volatile students were in the streets the past two days despite a blunt warning from the Pak government on Monday that the time had come for the students to get back to their books. The numbers of students involved have not been as large as in last month's outbursts, but the govern ment, which wants to get on with its negotiations with Japan, is clearly nettled. The police have been getting tougher and Pak has been thinking of imposing martial law if the students remain obstreperous. What makes the situation ticklish is the fact that the students are being egged on by opposition elements who want to bring down the government and by those whose main aim in life is to get rid of Kim Chong-pil. In short, all the ingredients are present for a nasty turn of events. (DOC_0005959127, 21 April 1964)

- Poland: As Edward Gierek prepares for his party's congress in December, he faces the most serious challenge to his political skills since the early months after he came to power. The volatile and demanding Polish people are increasingly restive over sporadic meat shortages and over prospective increases in food prices that they believe will lower their standard of living. The current leadership probably has both the means and the political acumen to avoid the mistakes that toppled the Gomulka regime in 1970, but as our embassy in Warsaw reports a "spark in the right place" could have serious consequences. (DOC_0006014933, 21 October 1975)

2.2.3 Predictive Process

To predict the likelihood of a racial trope appearing in a PDB entry, we apply a series of supervised learning models to our hand-coded trope data. We try five different models: multilayer perceptrons (MLP), C5.0 (decision trees), random forests (RF), Naive Bayes (NB), and gradient boosting (GBM).

Our PDB entry data, which is raw text, must be converted into a quantitative form to use these models. To that end, we translate our text data into six different forms. One is a classic bag-of-words approach where we produce a document-term matrix which indicates how many times a word/token j occurs in document/entry i.

However, we also create numerical representations of our text data using document embeddings, which are a particular manner in which a body of text (loosely called a "document") is converted into a single d-dimensional vector.[1] Document embeddings are created in a two-step process. First, word embeddings are generated for each word in a corpus. Word embeddings are a manner in which a word is represented as a vector of length d, which is calculated by analyzing the words surrounding a word in question. Two words that have similar word embeddings should be closer in meaning to one another. Second, a document embedding is constructed by "summing up" the word embeddings for all words in the document. As such, a document embedding based on word embeddings of length d will themselves also be of length d. Document embeddings account for the sequencing of words as well as some semantic features, unlike more traditional "bag of word" approaches that simply account for the number of times words are utilized in a text. This may be crucial to identifying concepts more subtly woven into language. Indeed, as we show in the appendix, the bag of words method fares poorly in predicting racial tropes. No single correct answer exists for how many dimensions should be used to represent a text. As our goal is to generate new measures using a predictive model, we simply use whichever representation of the data is most effective for each theme of interest. To test an array of possibilities, we convert our text into document embeddings of varying dimensions: 16, 32, 64, 128, and 256. (In other words, each PDB entry was translated into a vector of length 2^n, where n is an integer between 4 and 8.) These embeddings are plausibly more effective than bag-of-words approaches in capturing meanings and context across stretches of multiple words. To each of these representations, we can also choose whether to add information on word counts we obtain using our trope synonym dictionaries (see Appendix 2.1).

There are therefore six models for each of the five document embedding dimensions and the standard document-term matrix, six models for these same datasets with synonyms added, and one additional model using only synonyms. This is a total of 13 different representations of the text in the PDB entries. Each representation is used to train five separate supervised learning models. This means that each trope is predicted using 65 different combinations of data and models. Also note that within each of the supervised models, we also perform hyperparameter tuning to optimize validation performance, which means there are far more than 65 individual models underlying the prediction process.

We include two additional features across all models. The first is the presidential administration. This accounts for the well-established fact that different presidents read and requested information in the PDB in distinct ways (Priess, 2016, 95). Additionally, it is entirely plausible

[1] Further details about the algorithm, called Doc2Vec, are available in Le and Mikolov 2014.

Figure A2: Combinations of supervised learning methods. (This is a replication of Figure **??** in the main text, provided for convenience.)

Representation	**Add synonyms**	**Model**
$d = 16$	No	Multilayer perception
$d = 32$	Yes	C5.0 (Decision trees)
$d = 64$		Random forest
$d = 128$		Naive Bayes
$d = 256$		Gradient boosting
Bag of words		
None		

that presidents each harbor different views with respect to race (or their willingness to express such views), which could impact the degree to which PDBs reflect sentiments that align with their thinking. Second, we include a cubic spline for time to account for temporal interdependency and any potential time trends that may impact the prevalence of specific racial tropes.

2.2.4 Model Metrics

We create and assess the performance of models using our hand-coded data. A random sample of 70% of our hand-coded data is treated as a *training set*; the remaining 30% is held aside as a *test set*. Each of the 65 models attempts to "learn" the relationship between the hand-coded value for a trope and the raw quantitative data provided to represent the text of the associated PDB entry.

One perennial concern with training models is the risk of over-fitting. This concern is mitigated in two ways. First, the training step involves cross-validation, where the model is only trained on part of the training data, and then its predictive performance is assessed by seeing how well the model predicts the values for the remaining "validation set." The hyperparameters of a model are tuned through cross-validation. Second, once the best-performing version of the model is determined, it is applied to the test set. This produces predicted values for the test set (which was explicitly not used to originally train the model), and these predictions can be compared to the known actual values to assess the model's out-of-sample performance.

Several metrics could be used to assess a supervised model's performance. Four potential (and not comprehensive) candidates are Cohen's kappa coefficient, accuracy, F1 scores, and the area under the curve (AUC) statistic. Interested readers can find more in-depth definitions and discussions of each measure in a wide array of reference materials. For our current purposes, we note that accuracy (the proportion of times a model makes a correct prediction) is an especially poor metric for imbalanced data where there are relatively few positive cases. Cohen's Kappa, F1 scores, and AUC statistics are better-suited to deal with imbalanced data. Simulated experiments have shown that Kappa perhaps performs best (Fatourechi et al., 2008), which is why we use the Kappa coefficient to determine our best models.

Table A10 reports which combination of the PDB data and supervised models produced the best predictions in terms of kappa coefficients. A random forest model applied to a simple document-term matrix with counts of synonyms is most effective in discerning tropes related to irrationality. For the remaining three themes, a gradient boosting model offers best predictive performance. Note that animal and belligerence models work best when only provided information on the frequency of synonyms (without any data on token frequencies or document embeddings). This does not mean that these two models provide the same information as counting the frequency of synonyms. Our predictions of animal analogies and belligerence are based on hand-coded training data and thus reflect a level of judgment not present when using raw synonym counts measures of these tropes.

Table A10: Performance and descriptive statistics for highest-quality predictions.

	Representation	**Syns.**	**Model**	**Accuracy**	**Kappa**
Infantilization	Document embeddings ($d = 16$)	No	GBM	0.911	0.276
Animal analogy	None	Yes	GBM	0.952	0.560
Belligerence	None	Yes	GBM	0.848	0.542
Irrationality	Bag of words	Yes	RF	0.968	0.458

Full out-of-sample performance metrics for all models and tropes are presented in Tables A11 (Kappa), A12 (F1 score), A13 (AUC), and A14 (accuracy). The highest value for each trope, which indicates the best-performing model (reflected in Table A10, is bolded in each column.

Table A15 shows the confusion matrices for the test set based on the best-performing models for each racial trope.

Table A11: Kappa.

Text rep.	Syn.	Model	INF	ANIM	BELLIG	IRR
Bag of words	No	C5.0	0.020	0.059	0.072	0.092
Bag of words	No	GBM	-0.018	0.016	0.069	-0.003
Bag of words	No	MLP	0.006	0.015	0.048	-0.009
Bag of words	No	NB	0.000	0.000	0.018	0.000
Bag of words	No	RF	-0.003	0.119	0.094	0.016
Bag of words	Yes	C5.0	-0.010	0.404	0.338	0.429
Bag of words	Yes	GBM	-0.005	0.238	0.393	0.329
Bag of words	Yes	MLP	-0.013	0.000	0.046	0.029
Bag of words	Yes	NB	0.000	0.000	0.029	0.000
Bag of words	Yes	RF	-0.013	0.112	0.379	**0.458**
d=16	No	C5.0	-0.005	0.029	0.159	0.000
d=16	No	GBM	**0.276**	0.211	0.154	0.007
d=16	No	MLP	0.140	0.083	0.198	0.038
d=16	No	NB	0.122	0.132	0.121	0.012
d=16	No	RF	0.164	0.106	0.150	-0.007
d=16	Yes	C5.0	0.073	0.059	0.094	0.376
d=16	Yes	GBM	0.158	0.120	0.194	0.149
d=16	Yes	MLP	0.087	0.197	0.181	0.192
d=16	Yes	NB	0.150	0.031	0.191	-0.001
d=16	Yes	RF	0.115	0.195	0.115	0.283
d=32	No	C5.0	0.000	0.029	0.172	-0.007
d=32	No	GBM	0.173	0.225	0.278	0.072
d=32	No	MLP	0.213	0.178	0.256	0.062
d=32	No	NB	0.103	0.169	0.149	0.038
d=32	No	RF	0.117	0.056	0.211	-0.010
d=32	Yes	C5.0	0.073	0.000	0.065	0.352
d=32	Yes	GBM	0.125	0.152	0.276	0.151
d=32	Yes	MLP	0.051	0.233	0.202	0.136
d=32	Yes	NB	0.114	0.052	0.212	0.010
d=32	Yes	RF	0.098	0.047	0.218	0.352
d=64	No	C5.0	0.006	0.072	0.024	-0.004
d=64	No	GBM	0.177	0.291	0.289	0.059
d=64	No	MLP	0.120	0.059	0.260	0.063
d=64	No	NB	0.121	0.199	0.150	0.073
d=64	No	RF	0.111	0.142	0.246	0.000
d=64	Yes	C5.0	0.177	0.026	0.042	0.406
d=64	Yes	GBM	0.105	0.163	0.333	0.314
d=64	Yes	MLP	0.221	0.239	0.191	0.210
d=64	Yes	NB	0.155	0.079	0.198	0.014
d=64	Yes	RF	0.084	-0.018	0.105	0.058
d=128	No	C5.0	0.000	0.119	0.045	0.049
d=128	No	GBM	0.208	0.175	0.278	0.141
d=128	No	MLP	0.104	0.235	0.305	-0.022
d=128	No	NB	0.099	0.149	0.162	0.065
d=128	No	RF	0.089	0.119	0.179	0.000
d=128	Yes	C5.0	0.041	0.153	0.006	0.394
d=128	Yes	GBM	0.227	0.152	0.340	0.270
d=128	Yes	MLP	0.113	0.095	0.256	0.247
d=128	Yes	NB	0.132	0.112	0.225	0.021
d=128	Yes	RF	0.114	0.101	0.047	-0.004
d=256	No	C5.0	0.165	0.129	0.000	-0.007
d=256	No	GBM	0.231	0.189	0.293	0.151
d=256	No	MLP	0.173	0.143	0.275	-0.000
d=256	No	NB	0.090	0.127	0.143	0.036
d=256	No	RF	0.059	0.156	0.171	0.000
d=256	Yes	C5.0	-0.009	0.047	0.000	0.352
d=256	Yes	GBM	0.168	0.176	0.332	0.190
d=256	Yes	MLP	0.133	0.116	0.234	0.141
d=256	Yes	NB	0.153	0.101	0.189	0.033
d=256	Yes	RF	0.101	0.153	0.078	0.364
None	Yes	C5.0	-0.025	0.518	0.541	0.383
None	Yes	GBM	-0.005	**0.560**	**0.542**	0.195
None	Yes	MLP	0.000	0.050	0.011	0.000
None	Yes	NB	0.000	0.000	0.015	0.000
None	Yes	RF	-0.034	0.204	0.430	0.305

Table A12: F1.

Text rep.	Syn.	Model	INF	ANIM	BELLIG	IRR
Bag of words	No	C5.0	0.944	0.968	0.822	0.971
Bag of words	No	GBM	0.909	0.728	0.761	0.941
Bag of words	No	MLP	0.910	0.941	0.808	0.954
Bag of words	No	NB			0.284	
Bag of words	No	RF	0.927	0.944	0.859	0.963
Bag of words	Yes	C5.0	0.929	**0.975**	0.896	0.981
Bag of words	Yes	GBM	0.904	0.947	0.867	0.973
Bag of words	Yes	MLP	0.477	0.969	0.515	0.567
Bag of words	Yes	NB			0.248	
Bag of words	Yes	RF	0.910	0.956	0.879	**0.983**
d=16	No	C5.0	0.948	0.953	0.828	0.971
d=16	No	GBM	0.915	0.909	0.688	0.928
d=16	No	MLP	0.803	0.904	0.666	0.683
d=16	No	NB	0.786	0.726	0.526	0.775
d=16	No	RF	**0.953**	0.951	0.833	0.969
d=16	Yes	C5.0	0.949	0.950	0.861	0.979
d=16	Yes	GBM	0.901	0.861	0.788	0.926
d=16	Yes	MLP	0.658	0.924	0.594	0.946
d=16	Yes	NB	0.818	0.332	0.751	0.279
d=16	Yes	RF	0.943	0.955	0.848	0.978
d=32	No	C5.0	0.950	0.953	0.849	0.969
d=32	No	GBM	0.901	0.869	0.756	0.801
d=32	No	MLP	0.887	0.932	0.735	0.800
d=32	No	NB	0.743	0.753	0.553	0.791
d=32	No	RF	0.951	0.949	0.851	0.968
d=32	Yes	C5.0	0.949	0.960	0.870	0.978
d=32	Yes	GBM	0.909	0.873	0.771	0.906
d=32	Yes	MLP	0.904	0.951	0.795	0.882
d=32	Yes	NB	0.680	0.446	0.725	0.392
d=32	Yes	RF	0.939	0.959	0.863	0.978
d=64	No	C5.0	0.940	0.953	0.856	0.970
d=64	No	GBM	0.909	0.924	0.815	0.751
d=64	No	MLP	0.908	0.938	0.731	0.943
d=64	No	NB	0.765	0.777	0.564	0.749
d=64	No	RF	0.949	0.951	0.859	0.971
d=64	Yes	C5.0	0.945	0.953	0.875	0.979
d=64	Yes	GBM	0.807	0.919	0.820	0.968
d=64	Yes	MLP	0.917	0.936	0.814	0.960
d=64	Yes	NB	0.739	0.549	0.710	0.493
d=64	Yes	RF	0.944	0.955	0.856	0.973
d=128	No	C5.0	0.950	0.954	0.860	0.970
d=128	No	GBM	0.878	0.870	0.816	0.924
d=128	No	MLP	0.917	0.940	0.839	0.945
d=128	No	NB	0.686	0.761	0.635	0.717
d=128	No	RF	0.952	0.954	0.858	0.971
d=128	Yes	C5.0	0.951	0.961	0.870	0.978
d=128	Yes	GBM	0.886	0.890	0.841	0.938
d=128	Yes	MLP	0.905	0.940	0.823	0.966
d=128	Yes	NB	0.719	0.639	0.709	0.567
d=128	Yes	RF	0.951	0.960	0.854	0.972
d=256	No	C5.0	0.948	0.948	0.854	0.969
d=256	No	GBM	0.900	0.908	0.838	0.941
d=256	No	MLP	0.926	0.944	0.812	0.939
d=256	No	NB	0.698	0.745	0.578	0.716
d=256	No	RF	0.945	0.961	0.860	0.971
d=256	Yes	C5.0	0.948	0.959	0.872	0.978
d=256	Yes	GBM	0.884	0.861	0.855	0.934
d=256	Yes	MLP	0.920	0.946	0.809	0.964
d=256	Yes	NB	0.726	0.665	0.649	0.645
d=256	Yes	RF	0.948	0.961	0.859	0.979
None	Yes	C5.0	0.808	0.964	**0.904**	0.978
None	Yes	GBM	0.831	0.962	0.887	0.947
None	Yes	MLP	0.947	0.526	0.517	0.972
None	Yes	NB	0.947		0.174	
None	Yes	RF	0.814	0.950	0.864	0.970

Table A13: AUC.

Text rep.	**Syn.**	**Model**	**INF**	**ANIM**	**BELLIG**	**IRR**
Bag of words	No	C5.0	0.525	0.633	0.576	0.535
Bag of words	No	GBM	0.530	0.582	0.552	0.564
Bag of words	No	MLP	0.508	0.525	0.526	0.452
Bag of words	No	NB	0.500	0.500	0.518	0.500
Bag of words	No	RF	0.533	0.627	0.560	0.581
Bag of words	Yes	C5.0	0.571	0.844	0.778	0.677
Bag of words	Yes	GBM	0.574	0.773	0.764	0.706
Bag of words	Yes	MLP	0.485	0.607	0.536	0.610
Bag of words	Yes	NB	0.500	0.500	0.526	0.500
Bag of words	Yes	RF	0.539	0.790	0.770	0.600
d=16	No	C5.0	0.708	0.755	0.657	0.557
d=16	No	GBM	0.735	0.810	0.656	0.595
d=16	No	MLP	0.671	0.723	0.663	0.614
d=16	No	NB	0.682	0.785	0.642	0.599
d=16	No	RF	0.699	0.762	0.645	0.570
d=16	Yes	C5.0	0.710	0.684	0.693	0.706
d=16	Yes	GBM	0.687	0.717	0.699	0.682
d=16	Yes	MLP	0.663	0.723	0.696	0.745
d=16	Yes	NB	0.696	0.726	0.673	0.656
d=16	Yes	RF	0.711	0.724	0.683	0.690
d=32	No	C5.0	0.702	0.767	0.707	0.631
d=32	No	GBM	0.725	0.801	0.742	0.654
d=32	No	MLP	0.689	0.733	0.706	0.634
d=32	No	NB	0.675	0.798	0.661	0.644
d=32	No	RF	0.706	0.790	0.708	0.622
d=32	Yes	C5.0	0.721	0.704	0.716	0.701
d=32	Yes	GBM	0.701	0.721	0.744	0.726
d=32	Yes	MLP	0.646	0.643	0.677	0.717
d=32	Yes	NB	0.732	0.765	0.700	0.631
d=32	Yes	RF	0.730	0.742	0.733	0.699
d=64	No	C5.0	0.688	0.816	0.711	0.604
d=64	No	GBM	0.744	0.825	0.722	0.662
d=64	No	MLP	0.662	0.634	0.715	0.614
d=64	No	NB	0.704	0.825	0.671	0.686
d=64	No	RF	0.730	0.802	0.713	0.654
d=64	Yes	C5.0	0.659	0.707	0.732	0.655
d=64	Yes	GBM	0.689	0.690	0.745	0.729
d=64	Yes	MLP	0.700	0.702	0.663	0.716
d=64	Yes	NB	0.734	0.764	0.687	0.645
d=64	Yes	RF	0.714	0.718	0.743	0.696
d=128	No	C5.0	0.663	0.733	0.727	0.607
d=128	No	GBM	0.723	0.784	0.741	0.695
d=128	No	MLP	0.703	0.768	0.734	0.614
d=128	No	NB	0.692	0.802	0.688	0.689
d=128	No	RF	0.725	0.731	0.726	0.698
d=128	Yes	C5.0	**0.759**	0.723	0.705	0.705
d=128	Yes	GBM	0.753	0.753	0.758	**0.785**
d=128	Yes	MLP	0.657	0.724	0.730	0.684
d=128	Yes	NB	0.723	0.760	0.718	0.648
d=128	Yes	RF	0.730	0.721	0.731	0.739
d=256	No	C5.0	0.693	0.728	0.739	0.671
d=256	No	GBM	0.737	0.774	0.729	0.740
d=256	No	MLP	0.681	0.713	0.705	0.601
d=256	No	NB	0.697	0.787	0.660	0.670
d=256	No	RF	0.728	0.797	0.723	0.725
d=256	Yes	C5.0	0.723	0.725	0.732	0.648
d=256	Yes	GBM	0.741	0.763	0.754	0.711
d=256	Yes	MLP	0.671	0.714	0.712	0.704
d=256	Yes	NB	0.727	0.741	0.708	0.636
d=256	Yes	RF	0.747	0.734	0.751	0.687
None	Yes	C5.0	0.518	0.836	0.812	0.603
None	Yes	GBM	0.509	**0.870**	**0.819**	0.641
None	Yes	MLP	0.519	0.601	0.502	0.588
None	Yes	NB	0.527	0.500	0.785	0.500
None	Yes	RF	0.502	0.828	0.796	0.680

Table A14: Accuracy.

Text rep.	Syn.	Model	INF	ANIM	BELLIG	IRR
Bag of words	No	C5.0	0.894	0.938	0.712	0.944
Bag of words	No	GBM	0.834	0.585	0.642	0.889
Bag of words	No	MLP	0.836	0.888	0.694	0.912
Bag of words	No	NB	0.099	0.062	0.301	0.054
Bag of words	No	RF	0.864	0.895	0.762	0.930
Bag of words	Yes	C5.0	0.867	**0.952**	0.823	0.964
Bag of words	Yes	GBM	0.826	0.901	0.793	0.948
Bag of words	Yes	MLP	0.369	0.939	0.448	0.417
Bag of words	Yes	NB	0.128	0.061	0.314	0.045
Bag of words	Yes	RF	0.835	0.916	0.805	**0.968**
d=16	No	C5.0	0.901	0.910	0.726	0.944
d=16	No	GBM	0.850	0.839	0.596	0.866
d=16	No	MLP	0.690	0.829	0.590	0.536
d=16	No	NB	0.667	0.598	0.487	0.641
d=16	No	RF	**0.911**	0.908	0.730	0.940
d=16	Yes	C5.0	0.904	0.905	0.763	0.960
d=16	Yes	GBM	0.825	0.765	0.687	0.865
d=16	Yes	MLP	0.527	0.862	0.534	0.899
d=16	Yes	NB	0.709	0.258	0.652	0.199
d=16	Yes	RF	0.894	0.916	0.746	0.957
d=32	No	C5.0	0.904	0.910	0.751	0.940
d=32	No	GBM	0.825	0.780	0.672	0.679
d=32	No	MLP	0.805	0.875	0.652	0.677
d=32	No	NB	0.616	0.632	0.509	0.662
d=32	No	RF	0.906	0.903	0.757	0.939
d=32	Yes	C5.0	0.904	0.923	0.773	0.957
d=32	Yes	GBM	0.837	0.783	0.682	0.832
d=32	Yes	MLP	0.828	0.908	0.695	0.794
d=32	Yes	NB	0.552	0.340	0.633	0.276
d=32	Yes	RF	0.887	0.921	0.773	0.957
d=64	No	C5.0	0.887	0.910	0.751	0.942
d=64	No	GBM	0.837	0.864	0.726	0.616
d=64	No	MLP	0.835	0.885	0.649	0.894
d=64	No	NB	0.643	0.662	0.515	0.614
d=64	No	RF	0.904	0.908	0.769	0.944
d=64	Yes	C5.0	0.897	0.910	0.779	0.960
d=64	Yes	GBM	0.692	0.854	0.736	0.939
d=64	Yes	MLP	0.852	0.882	0.713	0.924
d=64	Yes	NB	0.618	0.425	0.619	0.354
d=64	Yes	RF	0.894	0.913	0.757	0.948
d=128	No	C5.0	0.904	0.913	0.757	0.942
d=128	No	GBM	0.793	0.780	0.726	0.861
d=128	No	MLP	0.850	0.890	0.753	0.895
d=128	No	NB	0.557	0.639	0.561	0.576
d=128	No	RF	0.909	0.913	0.763	0.944
d=128	Yes	C5.0	0.906	0.926	0.771	0.958
d=128	Yes	GBM	0.805	0.808	0.759	0.886
d=128	Yes	MLP	0.830	0.887	0.730	0.935
d=128	Yes	NB	0.594	0.509	0.623	0.419
d=128	Yes	RF	0.906	0.923	0.751	0.946
d=256	No	C5.0	0.901	0.903	0.746	0.940
d=256	No	GBM	0.825	0.836	0.751	0.890
d=256	No	MLP	0.865	0.895	0.722	0.886
d=256	No	NB	0.567	0.619	0.522	0.572
d=256	No	RF	0.897	0.926	0.765	0.944
d=256	Yes	C5.0	0.901	0.921	0.773	0.957
d=256	Yes	GBM	0.800	0.767	0.773	0.879
d=256	Yes	MLP	0.855	0.898	0.713	0.931
d=256	Yes	NB	0.603	0.532	0.571	0.495
d=256	Yes	RF	0.901	0.926	0.759	0.958
None	Yes	C5.0	0.686	0.934	**0.848**	0.957
None	Yes	GBM	0.717	0.931	0.830	0.901
None	Yes	MLP	0.899	0.405	0.448	0.946
None	Yes	NB	0.899	0.089	0.300	0.054
None	Yes	RF	0.693	0.906	0.793	0.942

Table A15: Confusion matrices for best performing models.

(a) Infantilization

		Actual	
		No	Yes
Predicted	No	328	22
	Yes	39	17

(b) Animal analogies

		Actual	
		No	Yes
Predicted	No	346	15
	Yes	12	20

(c) Belligerence

		Actual	
		No	Yes
Predicted	No	325	39
	Yes	44	79

(d) Irrationality

		Actual	
		No	Yes
Predicted	No	530	17
	Yes	1	8

2.2.5 False Positives and Negatives

Tables A16 and A17 feature examples of false positives and false negatives from the best-performing models for each trope. False positives refer to cases where the supervised model deems the entry to have a high likelihood of exhibiting a trope, while research assistants' manual qualitative coding concludes that the entry does not exhibit the trope. Negative negatives refer to cases where the supervised model deems the entry to have a low likelihood of exhibiting a trope, while the manual qualitative coding concludes that the entry does exhibit the trope.

To obtain examples of false positives, we identified all entries that were manually coded as not featuring the trope, and then we extracted five to ten entries from this subset that have the highest probability of featuring the trope according to the best performing supervised model. To obtain examples of false negatives, we identified all entries that were manually coded as featuring the trope, and then we extracted the five to ten entries from this subset that have the lowest probability of featuring the trope according to the best performing supervised model. The entries shown in Tables A16 and A17 represent one final example chosen from each cluster of candidate entries, chosen because they are most illustrative of potential issues our predictive models had in identifying racial tropes.

We first turn to false positives in Table A16. Across these examples, one common theme is that false positives appear to be triggered by the presence of individual words that could plausibly have a contextual connection with each trope, but a closer qualitative reading indicates that the trope is not present.

In the case of infantilization, note words like "hysterical" or "waking up" in the provided example. Another entry on Laos notes a "squabble" during a military action (DOC_0005995981, 15 October 1962). The implication of the word "squabble" is that the dispute at hand is trivial. While this could potentially be a sign of infantilization and perhaps other tropes, a human coder would not necessarily see this as a sign of actors being immature.

False positives for animal analogies arise when words synonymous with animals are used. In the example provided, the word "duck" is used in the sense of avoidance and not in terms of animal behavior. Human coders could note this distinction and ultimately coded this entry as having no animal analogies. However, because many positive examples of this trope in our training data do indeed feature animal names, our supervised model may erroneously identify synonyms as indicators of animal analogies.

False positives for belligerence often include words related to anger that, when qualitatively reviewed, do not actually reflect belligerence. The example in Table A16 features President Sadat being quite moderate in tone. However, the entry notes that "he had rejected the angry and emotional response." The use of these two words likely led the supervised model to treat them as high in belligerence, despite the fact that the entry says he rejected this attitude. Another entry on the Soviet Union notes that Brezhnev condemned US "aggression" (DOC_0005974103, 4 November 1967). This single word likely led the supervised model to identify belligerence, even though the entry discusses the Soviet leader's mention of another actor's aggression.

For irrationality, false positives are typically associated with discussions of rising tensions – whether military, political, or diplomatic. The example from the Dominican Republic describes a situation where "the political temperature is becoming dangerously hot." Although this phrase does suggest growing discontent, it does not satisfy our coding criteria (see Table A7) where an actor is being described as irrational, paranoid, or exhibiting flaws in their decision-making processes.

Another entry from Congo describes financial negotiations between the Congolese and Belgians, noting that the Belgians "probably will be civil" and that president Mobutu "will soon cool down" (DOC_0005973806, 17 May 1967). We consider this false positive to be a borderline case where the entry plausibly implies that the actors may not act rationally. Although the research assistant did not code this entry as indicating irrationality, it is encouraging and reasonable that the model picked up on this language.

Table A16: Examples of false positives from best performing supervised learning models.

Theme	Example
Infantilization	A. Brazil: Pro-Communist Miguel Arreas governor of Pernambuco state in the northeast, is reliably quoted as saying recently that "We can socialize Brazil and then detach it from the West without the Americans becoming hysterical, without their waking up to the fact, and without their intervening militarily if we do it slowly, gradually and quietly." The US Consul in Recife has reported his alarm at the pace of Communist activity in Pernambuco. (DOC_0005996317, 27 April 1963)
Animal analogy	7. Japan: Yesterday's B-52 crash in Okinawa is being exploited to the hilt by the Japanese press; From now on the government will be under even greater pressure to seek removal of the bombers and to stiffen Japan's position on the status of the bases after Okinawa reverts to Japan. The crash may produce some political fallout for Sato, even though he still seems a shoo-in in next week's elections for the presidency of his party. He has been trying to duck the whole reversion issue, but his two challengers have been pushing for tighter controls on US military activity in Okinawa. The newly elected Okinawan chief executive has reiterated that he will demand removal of the bombers. (DOC_0005976472, 20 November 1968)
Belligerence	EGYPT: In his speech to the nation on Saturday President Sadat set forth a policy of surprising moderation apparently designed to demonstrate his continued desire for peace. He also signaled some limits to his patience. Sadat deliberately played down militant themes. He said he had decided to extend the UN Emergency Force mandate and to reopen the Suez Canal by June because of his concern about the reaction of "the world." He said he had rejected the angry and emotional response to the breakdown of disengagement negotiations that most expected from him. He explained that he believed that Egypt could not be responsible for confronting the "international community" with a sudden crisis by not renewing the UN mandate when it expires on April 24. Using a similar rationale for reopening the canal, Sadat said that Egypt cannot deprive the "peoples of the world" of an important trade route when the canal had been closed through "no fault" of theirs. Sadat issued a warning, however, in both instances.... (DOC_0006014759, 31 March 1975)
Irrationality	5. Dominican Republic: The political temperature is becoming dangerously hot as the Dominican election campaign goes into its final week. There have already been sporadic incidents of violence recently and the charges and countercharges of fraud, corruption, and intimidation are growing more numerous with each passing day. In this situation, the chances for serious disturbances before 1 June are very real. (DOC_0005968325, 23 May 1966)

We now turn to false negatives in Table A17.

The vast majority of the false negatives related to infantilization, including the example provided, feature the words "bickering" or "dickering." Another entry on China notes that the "Chinese are dickering with the Australians" for wheat (DOC_0005973824, 27 May 1967). Research assistants were clearly drawn by these terms – which many would likely have found quite novel – and saw them as strong signals of the infantilization trope. The example in Table A17 is arguably the strongest example of a questionable false negative, according to our qualitative reading after the fact. In other cases, entries discuss negotiation proceedings, where terms like "dickering" are more likely to be used (since the term refers to bargaining). These entries are more justifiably deemed by the supervised model to not feature the infantilization trope. Those cases are arguably promising signs that the overall supervised model is making useful contextual distinctions.

False negatives for animal analogies always include words that are associated with animals but were not deemed as expressing animal analogies by the supervised model. Another entry on Japan describes upcoming elections as a "hornet's nest" (DOC_0005974329, 15 March 1968). It is unclear why these entries are being predicted not to feature animal analogies. Based on these results, dictionary-based measures may not only be sufficient but slightly preferable to the measures derived from supervised learning. We do note, however, that our main results are largely unaffected by whether we use count or probability measures of this trope.

With respect to belligerence, false negatives frequently involve discussions of interstate negotiations or domestic economic affairs. The example highlights rising protests against the Egyptian government's economic state. Another entry regarding Libya mentions that Qadhafi expressed "fury" over lack of progress in talks with Egypt (DOC_0006014840, 3 July 1975). Upon further qualitative review, all entries do express what we would deem belligerence. However, the broader subject matter of these entries – negotiations and domestic spending – are typically discussed in a much more dry manner in other entries on similar subjects (while the example discusses violent protests due to economic conditions, most entries about economic conditions would not indicate belligerence), leading the overall supervised model to generate a final prediction that the belligerence trope is not present.

Finally, for false negatives, the example provided from Indonesia is a clear instance of the irrationality trope; it discusses Sukarno as being nervous, irritable, and paranoid. This, in our estimation, is the most blatant false negative from our model. In most other cases, false negatives arise from more subtle entries where an actor makes completely unsubstantiated claims, which the human coder deemed to be based on an irrational, paranoid frame of mind. Another false negative entry on Morocco reports that the "Moroccans have so far provided no evidence to substantiate their charge that Algerian troops were involved in the fighting in Spanish Sahara last weekend" (DOC_0006015033, 18 February 1976). Although a relative strength of the supervised learning approach compared to the dictionary approach is that it can identify more subtle context, it would be extremely difficult for the supervised learning model to make accurate assessments on irrationality for examples such as these.

Table A17: Examples of false negatives from best performing supervised learning models.

Theme	**Example**
Infantilization	6. Venezuela: Bickering is on the rise in Caracas as the high command hunts for scapegoats for the Guyana insurrection fiasco. Along with the inauguration of a new president in March, this could produce a major shakeup in the military hierarchy. (DOC_0005976557, 11 January 1969)
Animal analogy	4. South Vietnam: A number of leading Khanh subordinates continue to growl about the way the general is running things. One of them, a corps commander in the northern part of the country, fears that South Vietnam is becoming "another Laos" with various political groups unable or unwilling to pull together even though the military situation is deteriorating. (DOC_0005959170, 12 May 1964)
Belligerence	[Egypt] Egyptian workers and students demonstrating violently in Cairo yesterday over economic grievances called for a return to Nasirist socialism. Prime Miniter Hijazi was the chief object of the protesters' wrath, but President Sadat was also criticized for failing to bring about promised economic recovery after the October 1973 war. Discontent over shortages and inflation has been on the rise since last summer and this dissatisfaction has been compounded by restlessness over the pace of progress in peace negotiations. Additional disturbances over a wide range of issues are possible and, although President Sadat now intends to move against leftist agitators, yesterday's demonstration could give these elements and student malcontents further impetus. (DOC_0006007907, 2 January 1975)
Irrationality	4. Indonesia: A. Pre-negotiation maneuvers with the Dutch continue in New York with little progress apparent. Both sides claim disappointment with the other's adamant stand. C. Embassay Djakarta reports Indonesian Communists and bloc representatives are vigorously exploiting Sukarno's present highly nervous and irritable mood to arouse his suspicions of Western intentions. (DOC_0005992165, 29 January 1962)

3 Quantitative Descriptive Statistics

Table A18 features descriptive statistics for all continuous variables used in the quantitative analysis. Table A19 supplies analogous statistics for all binary and categorical variables.

Tables A20 is a correlation matrix of the four predicted measures of racial tropes; Table A21 is a correlation matrix of the explanatory and control variables used in the main analysis.

Table A18: Descriptive statistics for continuous variables.

Variable	**Min.**	**1Q**	**Med.**	**Mean**	**3Q**	**Max.**
INF (Supervised)	0.003	0.069	0.169	0.241	0.365	0.954
ANIM (Supervised)	0.003	0.036	0.055	0.105	0.133	0.998
BELLIG (Supervised)	0.043	0.162	0.283	0.294	0.398	0.987
IRR (Supervised)	0.000	0.028	0.058	0.080	0.106	0.976
INF (Synonyms)	0.000	0.000	0.000	0.098	0.000	8.000
ANIM (Synonyms)	0.000	0.000	0.000	0.086	0.000	8.000
BELLIG (Synonyms)	0.000	0.000	0.000	0.567	1.000	22.000
IRR (Synonyms)	0.000	0.000	0.000	0.478	1.000	20.000
Years since indep. (logged)	0.000	0.000	0.000	1.218	2.773	4.094
Conflict	0.000	0.000	0.000	0.328	0.000	3.000
Personalism	0.000	0.000	1.000	0.756	1.000	2.000
Leader tenure (raw)	0.000	528.000	1,696.000	2,959.000	5,221.000	13,359.000
US trade	0.000	0.051	0.126	0.136	0.179	0.657
US military aid	0.000	0.000	0.000	6.095	15.629	22.982
Entry length	10.000	63.000	102.000	138.400	168.000	2,634.000

Table A19: Descriptive statistics for binary variables.

Variable	**No (0)**	**Yes (1)**
Global South	35,951	53,495
Democracy	71,202	18,244
Leader mentioned	74,986	14,460
US defense	67,615	21,831
Americas	8,048	81,398
Asia	30,766	58,680
Eastern Europe	11,341	78,105
Middle East/Northern Africa	17,008	72,438
Oceania	323	89,123
Sub-Saharan Africa	5,434	84,012
Western Europe	16,526	72,920

Figure A3 illustrates how our four racial tropes have changed in prevalence over the years. Tropes clearly do not undergo a secular decline over the course of our data.

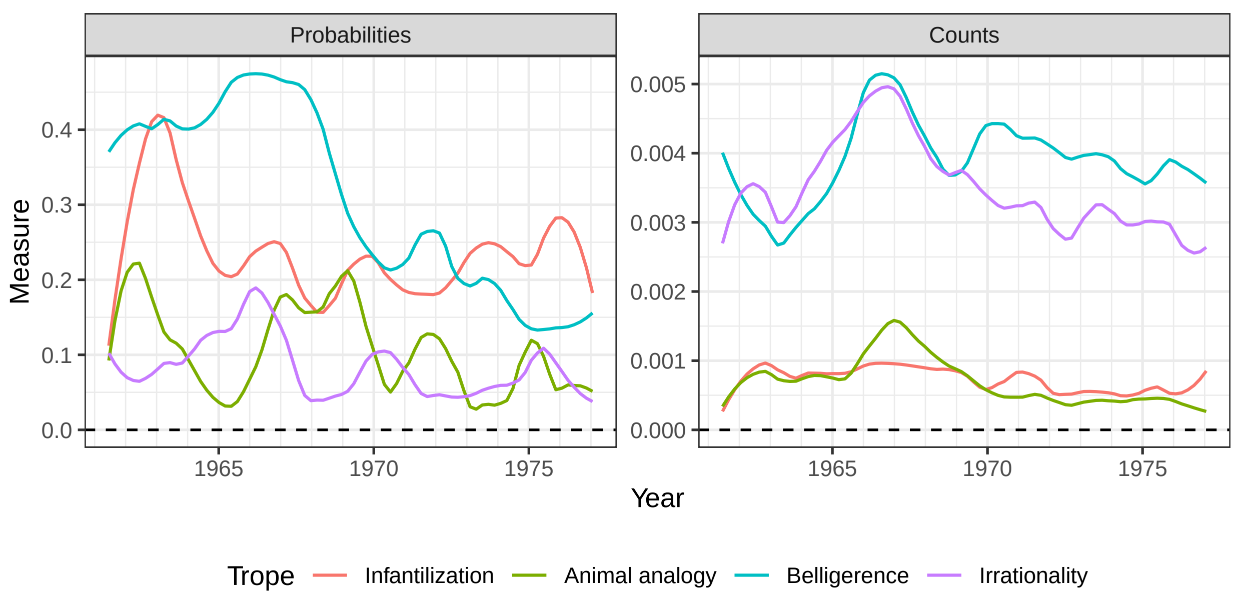
Table A20: Correlation matrix of outcome variables.

	Infant.	**Animal**	**Bellig.**	**Irrat.**
Infantilization	1.000	0.027	0.077	0.049
Animal analogy		1.000	0.130	-0.046
Belligerence			1.000	0.188
Irrationality				1.000

Table A21: Correlation matrix of explanatory and control variables.

	GS	**YSI**	**Conf.**	**Demo.**	**Pers.**	**Leader**	**Trade**	**Mil.**	**Def.**	**Words**
Global South	1.000	0.539	0.344	-0.457	0.299	-0.044	-0.449	0.015	-0.342	-0.031
Years since indep.		1.000	0.440	-0.173	-0.023	-0.025	-0.272	0.086	-0.396	0.071
Conflict			1.000	-0.210	0.019	-0.067	-0.334	-0.133	-0.224	0.028
Democracy				1.000	-0.512	0.067	0.615	0.184	0.525	0.070
Personalism					1.000	-0.026	-0.328	-0.150	-0.291	-0.030
Leader mention						1.000	0.079	0.113	0.121	0.094
US trade							1.000	0.244	0.653	0.036
US military aid								1.000	0.385	-0.041
US defense									1.000	-0.022
Entry length										1.000

Figure A3: Loess curves of trope prevalence over time. For purposes of standardization, counts are divided by total number of words in the entry.



4 Full Results and Robustness Checks

Below, we provide full results for our main analysis and robustness checks for our main results.

4.1 Full Results

Table A22 presents full results for models associated with Table 4 in the main text.

Table A22: Results of regressions on relationship between racial tropes in PDB entries and measurements of the racialized Otherness.

	Quasibinomial				*Poisson*			
	Infant.	Animal	Bellig.	Irrat.	Infant.	Animal	Bellig.	Irrat.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Global South	1.691^{***}	1.758^{***}	3.648^{***}	1.179^{**}	1.164^{***}	2.555^{***}	-0.066	0.955^{***}
	(0.322)	(0.514)	(0.922)	(0.518)	(0.321)	(0.747)	(0.150)	(0.226)
Years since independence	-0.492^{***}	-0.593^{***}	-1.040^{***}	-0.394^{***}	-0.483^{***}	-0.637^{***}	-0.030	-0.315^{***}
	(0.079)	(0.119)	(0.221)	(0.135)	(0.078)	(0.179)	(0.036)	(0.055)
Conflict	-0.256^{***}	0.133^{**}	0.0001	-0.231^{***}	-0.058^{*}	-0.025	-0.023	-0.022
	(0.055)	(0.056)	(0.038)	(0.036)	(0.030)	(0.041)	(0.019)	(0.019)
Democracy	-0.089	-0.103	-0.142	-0.176	-0.041	-0.015	0.035	-0.050
	(0.105)	(0.131)	(0.130)	(0.117)	(0.088)	(0.120)	(0.064)	(0.034)
Personalism	0.024	-0.002	-0.012	-0.046	-0.042	0.123	0.054^{**}	0.012
	(0.054)	(0.084)	(0.063)	(0.074)	(0.030)	(0.082)	(0.025)	(0.028)
Leader mention	0.378^{***}	-0.100^{***}	0.027	0.036^{*}	0.020	0.014	0.056^{***}	0.133^{***}
	(0.025)	(0.022)	(0.022)	(0.019)	(0.040)	(0.043)	(0.018)	(0.018)
US trade	-0.197	-2.781^{*}	-2.023	-0.138	-0.801	-2.318^{*}	-0.443^{*}	-0.812^{***}
	(0.851)	(1.632)	(2.504)	(1.102)	(0.573)	(1.230)	(0.263)	(0.266)
US military aid	0.023^{***}	0.010^{**}	0.033^{***}	0.009	-0.0001	0.012^{**}	-0.0004	0.007^{***}
	(0.003)	(0.004)	(0.006)	(0.009)	(0.004)	(0.006)	(0.002)	(0.002)
US defense	0.556^{***}	0.959^{***}	0.530^{**}	-0.263^{*}	-0.206	0.387^{**}	-0.088^{*}	0.149
	(0.088)	(0.164)	(0.260)	(0.135)	(0.161)	(0.155)	(0.052)	(0.122)
Entry length	0.331^{***}	-0.049^{***}	-0.107^{***}	0.129^{***}	1.060^{***}	0.989^{***}	1.127^{***}	1.064^{***}
	(0.012)	(0.015)	(0.017)	(0.021)	(0.024)	(0.020)	(0.010)	(0.008)
Constant	-2.510^{***}	-1.401^{***}	-0.184	-3.130^{***}	-7.037^{***}	-7.148^{***}	-6.071^{***}	-5.863^{***}
	(0.135)	(0.258)	(0.409)	(0.263)	(0.173)	(0.148)	(0.077)	(0.053)
Observations	89,016	89,016	89,016	89,016	89,016	89,016	89,016	89,016
Country FEs	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SEs (country)	✓	✓	✓	✓	✓	✓	✓	✓

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

4.2 Presidents and Time

In generating probabilistic measures of racial tropes, our supervised learning models include president fixed effects and a cubic time spline. These are included in the supervised models to account for any variation in racial tropes that could result from the interests or biases of the president, as well as broader societal changes that could affect when and how racialized language is deployed in public discourse. The quasibinomial models in our main findings, which use the final predictions from our best-performing supervised models, exhibit our predicted relationships.

Table A23 returns to the Poisson models in our main Global South analysis, but we now include president fixed effects and a cubic time spline into these models where the outcome variable does not account for the leader or broader time-based shifts. We see that our findings are not changed by these additions. Countries associated with the Global South face higher rates of racial tropes (though, as was the case in the main findings, the result for belligerence is not statistically significant), and countries that accumulate more years of independence are discussed with progressively fewer terms associated with racial tropes.

The coefficient estimates for presidents are worth some brief discussion. Note that Kennedy is used as the baseline category and therefore does not appear in the table. We see some evidence that the rate of some racial tropes appears to decline with more recent presidents. Compared to the Kennedy administration, Johnson's PDBs do not lean into as many analogies; Nixon's PDBs appear to feature fewer words associated with infantilization, animal analogies, and irrationality; and Ford's PDBs do not use terms tied to infantilization or animal analogies. Given ample anecdotal evidence about Johnson's and Nixon's racial bigotry, the fact that their intelligence briefings feature fewer racial tropes provides suggestive evidence that CIA officials did not adjust their language to simply mirror the perceived ideology of the leader.

Table A23: Results of regressions on relationship between racial tropes in PDB entries and measurements of the racialized Otherness, including president fixed effects and cubic time spline.

	Poisson			
	Infant.	Animal	Bellig.	Irrat.
	(1)	(2)	(3)	(4)
Global South	0.904^{***}	1.462^{***}	0.233	0.622^{***}
	(0.252)	(0.391)	(0.186)	(0.169)
Years since independence	-0.392^{***}	-0.252^{***}	-0.106^{**}	-0.194^{***}
	(0.064)	(0.095)	(0.047)	(0.042)
Conflict	-0.052^{*}	-0.092^{***}	-0.053^{***}	-0.059^{***}
	(0.028)	(0.024)	(0.020)	(0.012)
Democracy	−0.061	0.008	0.011	−0.032
	(0.087)	(0.138)	(0.063)	(0.033)
Personalism	-0.050^{*}	0.109^{***}	0.052^{***}	0.016
	(0.026)	(0.035)	(0.018)	(0.017)
Leader mention	0.022	0.036	0.058^{***}	0.137^{***}
	(0.040)	(0.045)	(0.018)	(0.016)
US trade	-1.005^{*}	−0.863	−0.310	−0.118
	(0.565)	(0.623)	(0.251)	(0.307)
US military aid	−0.002	0.001	0.003	0.005^{***}
	(0.005)	(0.004)	(0.002)	(0.002)
US defense	−0.097	0.152^{*}	0.072	0.142
	(0.160)	(0.081)	(0.049)	(0.099)
Entry length	1.092^{***}	1.106^{***}	1.118^{***}	1.094^{***}
	(0.027)	(0.017)	(0.008)	(0.008)
Johnson	-0.126^{*}	-0.306^{***}	0.034	0.033
	(0.076)	(0.079)	(0.059)	(0.040)
Nixon	-0.386^{***}	-1.114^{***}	−0.030	-0.166^{***}
	(0.097)	(0.112)	(0.062)	(0.046)
Ford	-0.244^{**}	-0.612^{***}	−0.005	−0.043
	(0.121)	(0.110)	(0.065)	(0.051)
Constant	-7.483^{***}	-8.196^{***}	-6.347^{***}	-6.363^{***}
	(0.152)	(0.162)	(0.079)	(0.083)
Observations	89,016	89,016	89,016	89,016
Country FEs	✓	✓	✓	✓
Cubic time spline	✓	✓	✓	✓
Clustered SEs (country)	✓	✓	✓	✓

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

4.3 Entry Topics with a Structural Topic Model

Our main statistical models in the paper do not directly account for the subject matter of a particular PDB entry besides the country or countries it involves. To ensure that our results are not driven by the subject matter of the PDB entries, we apply a structural topic model (STM) to our entry data Roberts, Stewart and Tingley (2017). We fit an STM that includes 65 topics. The number 65 was determined by using the algorithm outlined by Mimno and Lee (2014). This STM then generates topic propensities for each entry across all 65 topics. As such, each PDB entry is represented as a 65-dimensional probability vector (which adds up to 1).

Table A24 displays the 65 topics and distributional statistics for their propensities. The STM generates groups of words and documents that are deemed to exhibit a particular topic, but the model does not independently produce labels for each topic. We therefore reviewed the output of the STM – which include lists of words deemed to be indicative of each topic, as well as PDB entries which have high propensities in each topic – to create meaningful labels manually. Table A25 shows the ten words deemed to be most indicative of each topic in terms of FREX (a composite measure of a word's frequency and exclusivity for at topic). Note, however, that this word list is only one element of the several listed above that is used to determine a topic's label.

We performed two checks to validate the coherence of the topics which were produced by our 65-topic model, as well as the labels we created for each topic through qualitative review. To assess topic coherence, we use the Random 4 Word Set Intrusion task as outlined by Ying, Montgomery and Stewart (2022). We take a random topic and draw three random clusters of four words that the STM deems to have a high probability of being associated with the topic. Then, a random cluster of four high-probability words are drawn from a different randomly selected topic. This latter set of words constitute the intruder word set. The task for a human coder is to determine which of these four sets of words (after being randomly ordered) constitutes the intruder set. We created 500 of these tasks (and thus a total of 2,000 four-word sets) and qualitatively attempted to identify the intruder set in these 500 tasks. The overall accuracy of identifying the intruder set using our STM was 93.8%, which we consider to be quite high given that random guessing would lead to a rate of 25%. Table A26 reports the accuracy rates of identifying a topic as an intruder when it was the intruder topic in a specific task. Accuracy rates closer to 1 represent topics that are more easily distinguishable.

To assess the utility of our labels, we use the Optimal Labor task, also described by Ying, Montgomery, and Stewart.[2] A random PDB entry is drawn, and four topic labels are drawn with it. One of these labels is the correct label, while the others are not. A human coder is tasked with identifying which label is correct. We produced 500 of these tasks and then qualitatively attempted to identify the correct label for each entry. The overall accuracy for this task was 89.4%, which is again a respectable number given that random guesses would produce an accuracy of around 25%. Table A27 reports the accuracy of topic identification for each of the 65 topics.

Not all 65 topics were added to the final model. The 35 topics which are italicized in Table A24 were included as controls in our statistical analyses. In choosing what should be included as a potential control in our analysis, we wanted to only include topics that focused on particular forms of activity of concepts – particularly if they could bear any impact on the likelihood of racial tropes being used. While reviewing the most indicative words and representative texts for each topic, we

[2]In our version, we do not distinguish between the "within category" and "across categories" options.

noted two kinds of topics that would not be appropriate for inclusion. The first are topics that were synonymous with countries. Because our analyses already include country fixed effects, we do not want to add another variable that accounts for country. The second are headings for sections that appear on a near-daily basis across broad swaths of the PDB. These include so-called "topics" that reflect the existence of a table of contents, indicate the inclusion of late items, or reflect other common titles and non-substantive phrases used in the PDB. Topics in this latter category are flagged on Table Table A24 using quotation marks.

Table A28 displays the full results from regressions which include prevalence measures for the 35 topics we deemed to be relevant. As visually evident in Figure 3 of the main text, our primary findings are not affected.

For reference, Figure A4 shows the estimated coefficients for the 35 topic prevalence measures across all models reported in Table A28.

Table A24: Topics and propensities for PDB entries, according to 65-topic STM.

#	Topic	Min.	Q1	Med.	Mean	Q3	Max.
1	*Jewish migrants from USSR*	0.000	0.002	0.004	0.008	0.006	0.682
2	*Military buildups*	0.000	0.001	0.001	0.003	0.002	0.308
3	*Tensions*	0.000	0.002	0.003	0.007	0.006	0.309
4	*Indochina military activity*	0.000	0.002	0.004	0.026	0.013	0.850
5	*Domestic political difficulties*	0.000	0.003	0.005	0.008	0.008	0.272
6	*Naval aircraft and weapons*	0.000	0.003	0.007	0.022	0.014	0.771
7	"Major problems"	0.000	0.001	0.002	0.003	0.003	0.068
8	*Bilateral diplomatic relations*	0.000	0.002	0.004	0.009	0.007	0.337
9	*International negotiations*	0.000	0.008	0.015	0.034	0.030	0.773
10	China	0.000	0.002	0.003	0.015	0.008	0.566
11	*Trade and aid imports*	0.000	0.001	0.001	0.002	0.002	0.081
12	North Vietnam	0.000	0.001	0.003	0.022	0.014	0.615
13	*Ministerial politics*	0.000	0.007	0.013	0.029	0.026	0.738
14	*Communist messaging*	0.000	0.004	0.007	0.015	0.012	0.469
15	Berlin	0.000	0.001	0.002	0.012	0.004	0.731
16	Arab States	0.000	0.001	0.002	0.015	0.005	0.842
17	*Prisoner releases*	0.000	0.004	0.005	0.012	0.009	0.595
18	"Principal developments"	0.000	0.000	0.001	0.012	0.002	0.607
19	Cuba	0.000	0.001	0.002	0.011	0.005	0.659
20	*Coups*	0.000	0.006	0.012	0.039	0.032	0.786
21	*Rebel activities*	0.000	0.001	0.002	0.009	0.004	0.751
22	*Communist regime activities*	0.000	0.002	0.005	0.009	0.011	0.133
23	"Late items"	0.001	0.012	0.021	0.031	0.036	0.560
24	*Electoral politics*	0.000	0.003	0.006	0.020	0.012	0.817
25	Indonesia	0.000	0.002	0.005	0.019	0.010	0.833
26	*Oil/OPEC*	0.000	0.001	0.001	0.008	0.003	0.722
27	Spanish Sahara	0.000	0.001	0.002	0.007	0.003	0.827
28	*Third-party actions on Middle East*	0.000	0.003	0.005	0.014	0.009	0.714
29	India	0.000	0.002	0.003	0.012	0.005	0.690
30	Special Vietnam reports	0.000	0.003	0.004	0.007	0.007	0.183
31	West Berlin/Germany	0.000	0.001	0.002	0.004	0.003	0.450
32	Soviet Union	0.000	0.003	0.007	0.021	0.019	0.346
33	Indochina and Korea	0.000	0.002	0.004	0.021	0.012	0.878
34	Latin/South America	0.000	0.001	0.002	0.007	0.004	0.589
35	European community	0.000	0.003	0.004	0.015	0.008	0.884
36	"Developments"	0.000	0.001	0.001	0.003	0.002	0.147
37	"No significant changes"	0.000	0.003	0.004	0.007	0.007	0.233
38	*Socialist vs. democratic parties*	0.000	0.002	0.003	0.013	0.005	0.853
39	South Vietnam	0.000	0.002	0.004	0.017	0.011	0.553
40	Lebanon	0.000	0.001	0.001	0.010	0.002	0.864
41	*US attitudes to Vietnam War*	0.001	0.007	0.011	0.020	0.021	0.427
42	Laotian politics	0.000	0.001	0.002	0.012	0.004	0.862
43	*Asian trade*	0.000	0.001	0.002	0.008	0.004	0.686
44	*Diplomatic relations*	0.000	0.008	0.013	0.027	0.027	0.651
45	*Student demonstrations*	0.000	0.004	0.008	0.027	0.018	0.728
46	*Aircraft*	0.000	0.002	0.004	0.015	0.008	0.722
47	*Politburo affairs*	0.000	0.004	0.007	0.021	0.014	0.840
48	Warsaw Pact	0.000	0.002	0.003	0.016	0.007	0.838
49	Israel vs. Arab States	0.000	0.001	0.002	0.017	0.005	0.944
50	*Monetary affairs*	0.000	0.001	0.002	0.010	0.004	0.895
51	Combat in Laos	0.000	0.001	0.002	0.017	0.006	0.884
52	*Economic affairs*	0.000	0.004	0.007	0.028	0.014	0.926
53	*Negative expectations*	0.000	0.001	0.002	0.003	0.003	0.530
54	*Elections*	0.000	0.004	0.007	0.020	0.013	0.804
55	*Missiles and space*	0.000	0.003	0.006	0.033	0.014	0.963
56	Press regarding Vietnam	0.000	0.004	0.007	0.026	0.016	0.745
57	Congo	0.000	0.001	0.003	0.015	0.006	0.876
58	"Table of contents"	0.000	0.000	0.001	0.007	0.001	0.811
59	*Probabilistic assessments*	0.001	0.015	0.026	0.046	0.052	0.508
60	*Labor strikes*	0.000	0.002	0.003	0.010	0.006	0.664
61	*Military aid*	0.000	0.005	0.008	0.017	0.014	0.601
62	*Military strength/forces*	0.000	0.005	0.008	0.022	0.019	0.627
63	Greece vs. Turkey	0.000	0.001	0.002	0.013	0.003	0.862
64	*Popular movements in Africa*	0.000	0.002	0.003	0.012	0.005	0.794
65	Hong Kong and China	0.000	0.003	0.004	0.004	0.005	0.022

Table A25: Top ten words for each topic in terms of FREX.

#	Most indicative words (FREX)
1	document, emigr, appli, permiss, citizen, restrict, educ, sensit, expand, applic
2	note, con, discount, tinu, buildup, describ, briefli, full, slight, absent
3	situat, deterior, quiet, tens, calm, remain, worsen, wors, unchang, assess
4	regiment, kompong, phnom, penh, enemi, th, provinc, shell, mekong, rout
5	problem, difficulti, face, serious, solv, pose, solut, overcom, econom, resolv
6	union, naval, ship, water, port, sea, boat, mediterranean, submarin, vessel
7	major, affect, review, therebi, proport, settl, optimist, contribut, recognit, chiefli
8	interest, express, shown, privat, view, genuin, mutual, concern, dialogu, serv
9	negoti, agreement, propos, talk, panama, concess, treati, agre, progress, accept
10	china, chines, peke, chou, peip, taiwan, mao, en-lai, taipei, chiang
11	import, aim, of, compon, re, phase, breach, long-term, strengthen, pattern
12	north, vietnam, daili, vietnames, hanoi, dong, haiphong, reconstruct, parallel, raid
13	cabinet, minist, prime, resign, deputi, appoint, form, name, former, post
14	intern, reform, court, critic, world, detent, social, program, domest, nonalign
15	germani, berlin, brandt, bonn, german, pankow, chancellor, schmidt, east, ulbricht
16	sadat, cairo, arab, arabia, egypt, egyptian, husayn, plo, jordan, nasir
17	septemb, releas, prison, trial, sudan, numayri, return, boycott, sudanes, recal
18	princip, ii, page, discuss, cambodia, annex, cambodian, chile, appear, allend
19	america, havana, castro, cuba, cuban, mexican, mexico, venezuelan, venezuela, fidel
20	republ, coup, junta, plot, offic, armi, extremist, regim, assassin, arrest
21	dominican, bosch, balagu, garcia, somali, godoy, haiti, domingo, ethiopia, rebel
22	communist, communist-control, sign, upsurg, open, local, prelud, activ, tactic, build
23	morn, schedul, inform, late, night, yesterday, embassi, press, hour, announc
24	constitut, presidenti, allend, marco, peron, goulart, peronist, argentina, presid, amend
25	sukarno, indonesian, djakarta, indonesia, malaysia, zanzibar, british, london, guiana, jagan
26	saudi, shah, uar, iran, yemen, oil, kuwait, faysal, yemeni, opec
27	moroccan, morocco, algerian, spanish, bella, ben, sahara, franco, hassan, algeria
28	settlement, resolut, un, council, mandat, withdraw, peac, secur, secretari, debat
29	middl, pakistani, dacca, kashmir, indian, delhi, pakistan, bhutto, india, gandhi
30	report, special, brief, televis, news, mine, unconfirm, press, accord, correspond
31	west, establish, resid, ment, sought, occup, apprehens, offic, bank, let
32	soviet, moscow, gromyko, pravda, kosygin, russian, tass, brezhnev, ussr, podgorni
33	korea, korean, infiltr, cong, viet, truck, pyongyang, logist, bridg, southern
34	latin, american, peru, ecuador, velasco, uruguay, peruvian, oa, sanction, hemispher
35	french, ec, de, gaull, franc, common, nato, european, communiti, pari
36	develop, materi, transfer, raw, state, upcom, success, nairobi, advanc, automat
37	today, eastern, noth, affair, western, europ, overnight, chang, two-day, except
38	lon, sihanouk, socialist, democrat, nol, matak, coalit, christian, goncalv, cambodia
39	south, buddhist, ky, thieu, saigon, khanh, nang, da, huong, quang
40	lebanes, muslim, beirut, palestinian, christian, pak, jumblatt, syrian, lebanon, franjiyah
41	us, war, attitud, reflect, alleg, charg, aggress, polit, warn, violat
42	phoumi, souvanna, vientian, hong, pathet, neutralist, souphannouvong, lao, icc, phouma
43	japanes, japan, tokyo, trade, sato, tanaka, canada, canadian, credit, netherland
44	relat, diplomat, visit, trip, sino-soviet, thai, disput, asia, improv, thailand
45	student, demonstr, protest, polic, violenc, riot, disturb, agit, univers, street
46	fighter, aircraft, libyan, pilot, mig-, equip, iraqi, air, baghdad, jet
47	committe, leadership, politburo, congress, teng, parti, watch, chairman, revolut, central
48	czechoslovakia, warsaw, pact, romania, pragu, rumanian, rumania, khrushchev, bucharest, dubcek
49	isra, fedayeen, canal, aviv, israel, suez, tel, raid, syrian, golan
50	market, dollar, compani, currenc, monetari, float, fund, invest, exchang, bank
51	irregular, tieng, plain, pao, vang, ban, muong, boloven, jarr, des
52	product, million, ton, grain, crop, harvest, food, rice, plant, wheat
53	signific, confin, routin, thin, hit, scale, shop, wear, tire, sudden
54	vote, elect, seat, parliamentari, sunday, hous, win, poll, candid, assembl
55	space, photographi, test, construct, ss-, missil, silo, satellit, icbm, site
56	broadcast, articl, commentari, statement, comment, johnson, bomb, quot, radio, interview
57	tshomb, adoula, leopoldvill, congo, congoles, katanga, katangan, stanleyvill, mercenari, brazzavill
58	content, tabl, page, portug, lebanon, egypt-israel, annex, ussr, thailand, ethiopia
59	even, certain, cours, greater, longer, becom, feel, real, less, can
60	labor, britain, worker, wage, wilson, strike, membership, cut, unemploy, budget
61	aid, assist, advis, nigeria, request, help, feder, militari, technic, presenc
62	unit, command, region, divis, combat, evacu, forc, armi, strength, troop
63	cypriot, greek, makario, turkish, athen, nicosia, turk, ankara, clerid, cyprus
64	angola, movement, portugues, rhodesian, african, rhodesia, africa, popular, smith, zambia
65	kong, plan, continu, may, week, sign, anoth, seem, still, come

Table A26: Accuracy in Random 4 Word Set Intrusion task for individual topics in 65-topic STM.

#	Topic	Total	Correct	Accuracy
1	Jewish migrants from USSR	12	11	0.917
2	Military buildups	19	17	0.895
3	Tensions	16	16	1.000
4	Indochina military activity	10	9	0.900
5	Domestic political difficulties	11	10	0.909
6	Naval aircraft and weapons	13	13	1.000
7	"Major problems"	13	11	0.846
8	Bilateral diplomatic relations	16	16	1.000
9	International negotiations	20	20	1.000
10	China	13	13	1.000
11	Trade and aid imports	9	9	1.000
12	North Vietnam	8	8	1.000
13	Ministerial politics	14	14	1.000
14	Communist messaging	16	12	0.750
15	Berlin	6	6	1.000
16	Arab states	22	22	1.000
17	Prisoner releases	19	19	1.000
18	"Principal developments"	19	14	0.737
19	Cuba	14	14	1.000
20	Coups	15	14	0.933
21	Rebel activities	21	20	0.952
22	Communist regime activities	23	21	0.913
23	"Late items"	11	10	0.909
24	Electoral politics	9	9	1.000
25	Indonesia	15	13	0.867
26	Oil/OPEC	13	13	1.000
27	Spanish Sahara	13	13	1.000
28	Third-party actions on Middle East	12	12	1.000
29	India	17	16	0.941
30	Special daily reports	13	11	0.846
31	West Berlin/Germany	12	12	1.000
32	Soviet Union	15	15	1.000
33	Indochina and Korea	17	17	1.000
34	Latin/South America	18	16	0.889
35	European community	16	15	0.938
36	"Developments"	22	17	0.773
37	"No significant changes"	18	9	0.500
38	Socialist vs. democratic parties	18	18	1.000
39	South Vietnam	25	24	0.960
40	Lebanon	10	10	1.000
41	US attitudes to Vietnam War	8	6	0.750
42	Laotian politics	18	18	1.000
43	Asian trade	10	10	1.000
44	Diplomatic relations	27	27	1.000
45	Student demonstrations	13	13	1.000
46	Aircraft	17	17	1.000
47	Politburo affairs	18	18	1.000
48	Warsaw Pact	19	19	1.000
49	Israel vs. Arab states	17	16	0.941
50	Monetary affairs	15	15	1.000
51	Combat in Laos	22	22	1.000
52	Economic affairs	28	28	1.000
53	Negative expectations	21	19	0.905
54	Elections	12	12	1.000
55	Missiles and space	13	13	1.000
56	Press regarding Vietnam	14	14	1.000
57	Congo	10	10	1.000
58	"Table of contents"	22	16	0.727
59	Probabilistic assessments	10	9	0.900
60	Labor strikes	16	15	0.938
61	Military aid	7	6	0.857
62	Military strength/forces	10	10	1.000
63	Greece vs. Turkey	16	16	1.000
64	Popular movements in Africa	16	15	0.938
65	Hong Kong and China	18	15	0.833
	Overall tasks	500	469	0.938

Table A27: Accuracy in Optimal Labor Task for individual topics in 65-topic STM.

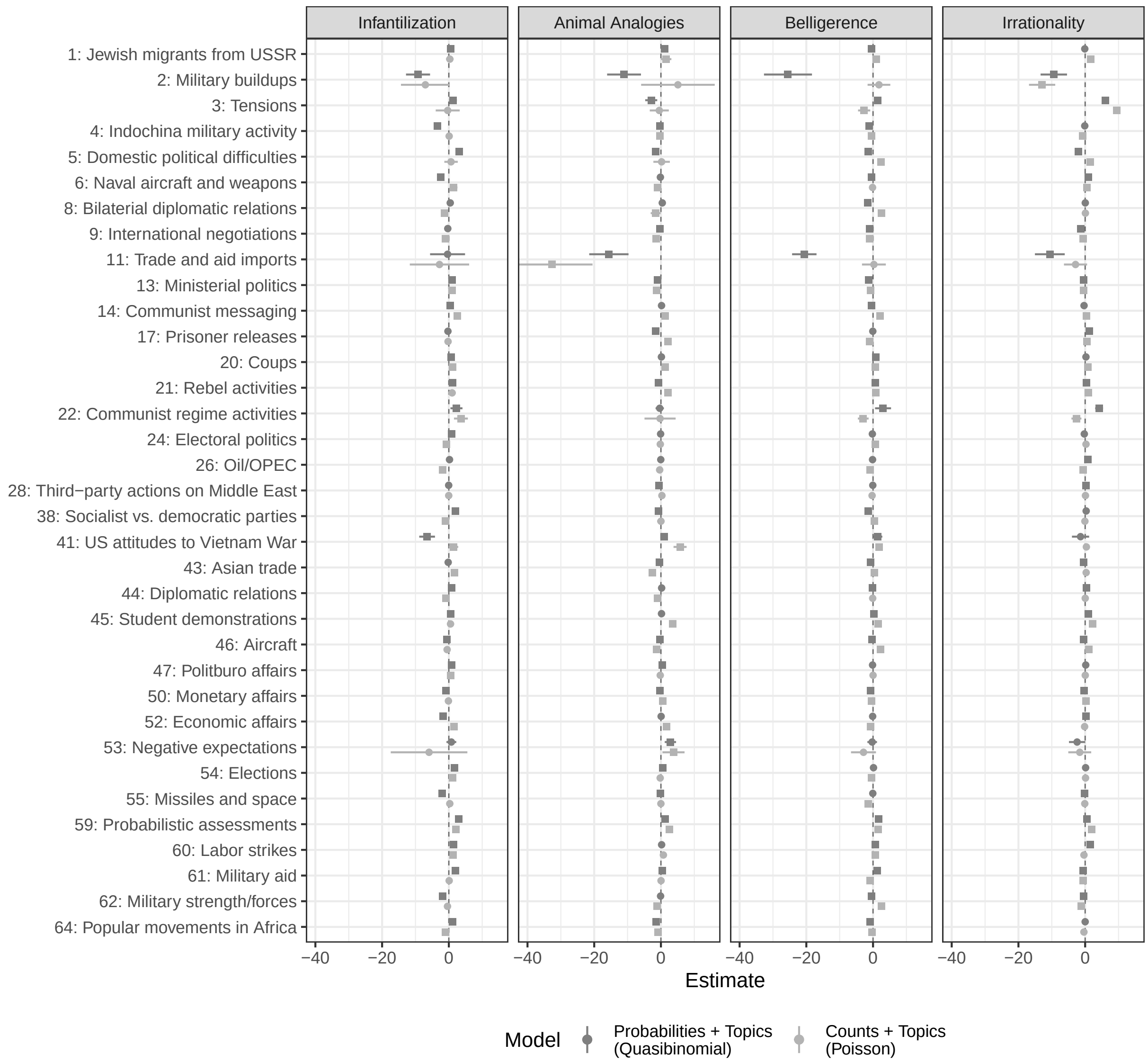
#	Topic	Total	Correct	Accuracy
1	Jewish migrants from USSR	29	25	0.862
2	Military buildups	13	12	0.923
3	Tensions	31	27	0.871
4	Indochina military activity	33	30	0.909
5	Domestic political difficulties	24	19	0.792
6	Naval aircraft and weapons	25	19	0.760
7	"Major problems"	20	19	0.950
8	Bilateral diplomatic relations	21	18	0.857
9	International negotiations	46	43	0.935
10	China	28	27	0.964
11	Trade and aid imports	22	20	0.909
12	North Vietnam	26	24	0.923
13	Ministerial politics	29	26	0.897
14	Communist messaging	29	24	0.828
15	Berlin	31	26	0.839
16	Arab states	41	41	1.000
17	Prisoner releases	28	25	0.893
18	Principal developments	30	29	0.967
19	Cuba	22	16	0.727
20	Coups	46	43	0.935
21	Rebel activities	40	32	0.800
22	Communist regime activities	20	15	0.750
23	"Late items"	26	25	0.962
24	Electoral politics	43	35	0.814
25	Indonesia	42	36	0.857
26	Oil/OPEC	25	23	0.920
27	Spanish Sahara	35	34	0.971
28	Third-party actions on Middle East	42	40	0.952
29	India	39	35	0.897
30	Special Vietnam reports	26	25	0.962
31	West Berlin/Germany	16	15	0.938
32	Soviet Union	18	16	0.889
33	Indochina and Korea	33	28	0.848
34	Latin/South America	25	20	0.800
35	European community	35	29	0.829
36	Developments	21	20	0.952
37	No significant changes	18	17	0.944
38	Socialist vs. democratic parties	29	27	0.931
39	South Vietnam	33	32	0.970
40	Lebanon	32	32	1.000
41	US attitudes to Vietnam War	23	21	0.913
42	Laotian politics	31	29	0.935
43	Asian trade	31	28	0.903
44	Diplomatic relations	37	34	0.919
45	Student demonstrations	36	33	0.917
46	Aircraft	29	27	0.931
47	Politburo affairs	32	27	0.844
48	Warsaw Pact	36	31	0.861
49	Israel vs. Arab states	27	24	0.889
50	Monetary affairs	20	17	0.850
51	Combat in Laos	38	36	0.947
52	Economic affairs	44	37	0.841
53	Negative expectations	31	29	0.935
54	Elections	48	46	0.958
55	Missiles and space	33	32	0.970
56	Press regarding Vietnam	40	35	0.875
57	Congo	27	25	0.926
58	"Table of contents"	37	36	0.973
59	Probabilistic assessments	36	27	0.750
60	Labor strikes	30	26	0.867
61	Military aid	25	21	0.840
62	Military strength/forces	40	34	0.850
63	Greece vs. Turkey	35	33	0.943
64	Popular movements in Africa	41	36	0.878
65	Hong Kong and China	21	15	0.714
	Overall tasks	500	447	0.894

Table A28: Results of regressions on relationship between racial tropes in PDB entries and measurements of the racialized Otherness, including topics from STM.

	Quasibinomial				*Poisson*			
	Infant.	Animal	Bellig.	Irrat.	Infant.	Animal	Bellig.	Irrat.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Global South	0.871***	1.410***	2.922***	1.087**	0.963***	1.920***	−0.265*	0.298**
	(0.258)	(0.462)	(0.789)	(0.480)	(0.268)	(0.595)	(0.140)	(0.122)
Years since independence	−0.291***	−0.491***	−0.835***	−0.362***	−0.381***	−0.424***	0.055	−0.130***
	(0.060)	(0.108)	(0.189)	(0.127)	(0.068)	(0.143)	(0.033)	(0.031)
Conflict	−0.169***	0.119**	−0.012	−0.219***	−0.045*	−0.004	0.005	0.029**
	(0.053)	(0.054)	(0.034)	(0.029)	(0.026)	(0.031)	(0.019)	(0.015)
Democracy	−0.077	−0.097	−0.115	−0.152	−0.033	−0.034	0.037	−0.038
	(0.098)	(0.129)	(0.110)	(0.111)	(0.085)	(0.114)	(0.060)	(0.036)
Personalism	0.009	0.001	−0.012	−0.041	−0.040	0.104	0.049	0.0005
	(0.043)	(0.078)	(0.052)	(0.068)	(0.029)	(0.067)	(0.032)	(0.025)
Leader mention	0.128***	−0.107***	0.017	0.024	0.006	0.007	0.033***	0.098***
	(0.016)	(0.022)	(0.018)	(0.018)	(0.034)	(0.036)	(0.012)	(0.013)
US trade	−0.389	−2.641*	−1.588	−0.003	−0.719	−2.274**	−0.635**	−0.850***
	(0.689)	(1.443)	(2.123)	(1.024)	(0.523)	(1.094)	(0.283)	(0.207)
US military aid	0.016***	0.007*	0.028***	0.008	−0.002	0.008	−0.005**	0.003*
	(0.004)	(0.004)	(0.005)	(0.009)	(0.005)	(0.005)	(0.002)	(0.002)
US defense	0.549***	0.891***	0.418*	−0.243*	−0.186	0.368***	−0.046	0.125
	(0.067)	(0.141)	(0.216)	(0.124)	(0.138)	(0.132)	(0.042)	(0.091)
Entry length	0.289***	−0.066***	−0.110***	0.134***	0.989***	0.952***	1.039***	1.019***
	(0.013)	(0.018)	(0.020)	(0.017)	(0.024)	(0.029)	(0.008)	(0.013)
Constant	−2.371***	−1.305***	−0.074	−3.157***	−7.095***	−7.198***	−5.891***	−5.791***
	(0.169)	(0.244)	(0.358)	(0.251)	(0.127)	(0.175)	(0.086)	(0.086)
Observations	89,016	89,016	89,016	89,016	89,016	89,016	89,016	89,016
Topics	✓	✓	✓	✓	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SEs (country)	✓	✓	✓	✓	✓	✓	✓	✓

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Figure A4: Coefficient plots for topics in regressions from Table A28. Points with squares represent statistical significance at the 95% level.



4.4 Including Leader Tenure

One may be concerned that the negative coefficients we find for our *years since independence* variable are actually a reflection of years that individual leaders remain in power. While that result may be interesting in its own right, it would also detract from our primary argument.

We address this by constructing a measure of *leader tenure* using information from the Archigos dataset (Goemans, Gleditsch and Chiozza, 2009). For each entry, we identify the head of state in the country at the time and calculate how many days they have held power. We log this value in our analysis. Note that this value is added to the model regardless of whether a leader is explicitly mentioned in the entry or not.

Table A29 replicates our main analysis after adding the measure of leader tenure. The inclusion of this variable results in the loss of over 7,000 observations where it was not possible to make a clean match between an entry and leader. (For example, an entry about "Germany" may not specify whether it is talking about East Germany or West Germany, which precludes our ability to measure leader tenure.) Nonetheless, the results are largely unchanged relative to the main manuscript.

Table A29: Results of regressions on relationship between racial tropes in PDB entries and measurements of the racialized Otherness, including leader tenure.

	Quasibinomial				*Poisson*			
	Infant.	Animal	Bellig.	Irrat.	Infant.	Animal	Bellig.	Irrat.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Global South	0.980^{***}	1.573^{***}	3.343^{***}	1.169^{**}	1.318^{***}	2.053^{***}	-0.296^{**}	0.301^{**}
	(0.279)	(0.475)	(0.696)	(0.500)	(0.242)	(0.544)	(0.144)	(0.129)
Years since indep.	-0.332^{***}	-0.489^{***}	-0.913^{***}	-0.403^{***}	-0.455^{***}	-0.466^{***}	0.056	-0.125^{***}
	(0.067)	(0.114)	(0.162)	(0.131)	(0.061)	(0.126)	(0.035)	(0.032)
Conflict	-0.171^{***}	0.122^{**}	0.014	-0.221^{***}	−0.044	0.008	0.006	0.031^{**}
	(0.053)	(0.056)	(0.031)	(0.030)	(0.027)	(0.031)	(0.019)	(0.015)
Democracy	−0.077	−0.077	−0.132	−0.157	−0.031	−0.050	0.042	−0.018
	(0.098)	(0.135)	(0.105)	(0.112)	(0.084)	(0.113)	(0.061)	(0.039)
Personalism	0.009	−0.022	−0.042	−0.028	−0.046	0.080	0.046	−0.0001
	(0.045)	(0.078)	(0.050)	(0.066)	(0.029)	(0.067)	(0.034)	(0.027)
Leader mention	0.132^{***}	-0.072^{***}	0.035^{*}	−0.007	0.017	0.016	0.035^{***}	0.106^{***}
	(0.016)	(0.022)	(0.019)	(0.019)	(0.033)	(0.037)	(0.012)	(0.014)
Leader tenure	-0.016^{***}	0.048^{***}	0.055^{***}	-0.041^{***}	0.0003	0.036^{***}	0.007^{*}	0.009^{*}
	(0.005)	(0.011)	(0.008)	(0.007)	(0.008)	(0.008)	(0.004)	(0.004)
US trade	−0.409	-2.606^{*}	−1.882	−0.009	-1.022^{**}	-2.692^{***}	-0.714^{**}	-0.687^{***}
	(0.709)	(1.447)	(2.058)	(1.067)	(0.488)	(1.025)	(0.287)	(0.231)
US mil. aid	0.016^{***}	0.006	0.027^{***}	0.008	0.001	0.007	-0.005^{**}	0.003^{*}
	(0.004)	(0.004)	(0.005)	(0.009)	(0.005)	(0.005)	(0.002)	(0.002)
US defense	0.542^{***}	0.896^{***}	0.461^{**}	-0.285^{**}	−0.143	0.397^{***}	−0.047	0.086
	(0.068)	(0.140)	(0.202)	(0.114)	(0.138)	(0.122)	(0.036)	(0.078)
Entry length	0.287^{***}	-0.056^{***}	-0.094^{***}	0.125^{***}	1.019^{***}	0.966^{***}	1.047^{***}	1.018^{***}
	(0.014)	(0.020)	(0.022)	(0.018)	(0.026)	(0.031)	(0.009)	(0.014)
Constant	-2.233^{***}	-1.803^{***}	−0.533	-2.810^{***}	-7.320^{***}	-7.392^{***}	-5.911^{***}	-5.874^{***}
	(0.173)	(0.253)	(0.375)	(0.275)	(0.153)	(0.209)	(0.099)	(0.106)
Observations	81,828	81,828	81,828	81,828	81,828	81,828	81,828	81,828
Topics	✓	✓	✓	✓	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SEs (country)	✓	✓	✓	✓	✓	✓	✓	✓

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

4.5 Only Decolonized States

Our measure of years since independence only accounts for instances of independence that took place following World War I. This was done to ensure that countries like Portugal – which was founded and recognized in 1143 CE – are not artificially inflating the measure. In cases where countries became independent before World War I, we coded *years since independence* as zero. This choice, however, may introduce another form of bias in the other direction.

The cleanest test of our argument regarding years since independence would only analyze countries that were formally decolonized in the post-World War I era. As predominantly non-Western countries gained sovereignty from Western powers that would have likely seen them through a negative and racialized lens, Western states' views should gradually improve as these newer states demonstrated their "maturity."

We perform this cleanest test in Table A30. We limit our data to only PDB entries that mention countries which were formally decolonized, as coded in the Issue Correlates of War Project's Colonial History Data Set (Hensel and Mitchell, 2007). This results in 39,598 units of analysis, which is about 44% of our overall data. Our findings are unchanged for the quasibinomial models which use our predicted measures of racial tropes. At the same time, the previously significant results for infantilization and animal analogies are neutralized in the Poisson models; the negative coefficient remains statistically significant for irrationality. These contrasting findings suggest that more complex and subtle forms of racial tropes are strongly connected to the perceived maturity of newly decolonized states, while simpler manifestations of racial tropes reflected by single words are not.

The vast majority of our models sustain the main finding: as decolonized states become more distant from their moment of independence, PDB entries mentioning these countries become less likely to rely on our racial tropes of interest.

Table A30: Results of regressions on relationship between racial tropes in PDB entries and measurements of the racialized Otherness, limited only to decolonized states.

	Dependent variable:							
	Quasibinomial				*Poisson*			
	Infant.	Animal	Bellig.	Irrat.	Infant.	Animal	Bellig.	Irrat.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years since independence	-0.411^{***}	-0.619^{***}	-1.077^{***}	-0.452^{***}	-0.362^{***}	-0.484^{***}	0.136^{***}	-0.116^{**}
	(0.059)	(0.103)	(0.176)	(0.147)	(0.079)	(0.158)	(0.040)	(0.050)
Conflict	-0.167^{***}	0.113^{*}	0.024	-0.193^{***}	-0.038^{*}	0.005	−0.008	0.029
	(0.057)	(0.060)	(0.028)	(0.032)	(0.020)	(0.039)	(0.016)	(0.017)
Democracy	0.098	−0.022	−0.011	−0.162	0.060	0.222	−0.090	−0.079
	(0.140)	(0.210)	(0.095)	(0.157)	(0.082)	(0.184)	(0.087)	(0.059)
Personalism	0.067	−0.087	−0.024	−0.024	-0.047^{*}	0.019	−0.002	−0.044
	(0.063)	(0.090)	(0.064)	(0.090)	(0.026)	(0.076)	(0.052)	(0.035)
Leader mention	0.157^{***}	-0.160^{***}	−0.008	0.002	0.069	0.025	0.025^{*}	0.100^{***}
	(0.026)	(0.033)	(0.024)	(0.034)	(0.044)	(0.057)	(0.014)	(0.022)
US trade	1.989^{**}	1.018	4.936^{***}	2.494	0.078	2.078	-1.219^{*}	-0.966^{**}
	(0.778)	(1.113)	(1.394)	(1.639)	(1.132)	(1.422)	(0.711)	(0.480)
US military aid	0.0001	0.002	−0.0001	−0.014	0.001	−0.006	0.002	0.004
	(0.004)	(0.006)	(0.009)	(0.013)	(0.007)	(0.007)	(0.002)	(0.002)
US defense	0.347	0.923^{***}	0.773^{***}	0.221	6.374^{***}	0.562	0.429^{*}	1.889
	(0.300)	(0.289)	(0.297)	(0.308)	(0.895)	(0.754)	(0.246)	(1.158)
Entry length	0.322^{***}	−0.006	−0.0002	0.161^{***}	0.959^{***}	0.937^{***}	1.025^{***}	1.022^{***}
	(0.029)	(0.027)	(0.021)	(0.025)	(0.052)	(0.062)	(0.010)	(0.026)
Constant	-1.428^{***}	0.238	3.167^{***}	-1.847^{***}	-6.041^{***}	-4.866^{***}	-6.319^{***}	-5.577^{***}
	(0.217)	(0.340)	(0.630)	(0.399)	(0.232)	(0.436)	(0.161)	(0.150)
Observations	39,598	39,598	39,598	39,598	39,598	39,598	39,598	39,598
Topics	✓	✓	✓	✓	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SEs (country)	✓	✓	✓	✓	✓	✓	✓	✓

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

4.6 More on Regional Variation

Table A31 presents the full results of regressions that split apart countries by geographical regions as opposed to the broad binary of whether they are considered part of the Global South. The main findings for regions are visually communicated in Figure 5 of the main text.

At least three countries are worth further investigation because of their unique attributes, particularly during the time period covered by our PDB data. These are Vietnam, the Soviet Union, and South Africa. Vietnam is substantively important due to the ongoing war during at least a decade of our data. It is also the most frequently discussed country in the PDB: 14% of all observations in our analysis involve Vietnam, and 42% of all entries about Asia are actually about Vietnam. As such, one may wonder whether the results we obtain regarding Asia are simply an artifact of Vietnam.

Similarly, the Soviet Union is the rival superpower during the Cold War. Entries on the USSR comprise 10% of our overall data and 80% of all entries regarding Eastern Europe. Indeed, its placement in "Eastern Europe" could be highly contested in public discussion, as some would instead consider it part of "Asia" or "Europe." We may therefore be curious to see how much of any effects we find for Eastern Europe are truly about the Soviet Union.

South Africa is not discussed frequently in the PDB. Only 259 entries (0.3%) involve South Africa, and South Africa constitutes less than 5% of entries regarding Sub-Saharan Africa. Nonetheless, South Africa under the racist rule of apartheid during this time was majority Black with a government dominated by a White minority. Clustering South Africa with the rest of Sub-Saharan Africa may underestimate our results.

Table A32 replicates the analysis in Table A31, but we split apart these three countries from their regions and estimate their effects separately. With respect to Vietnam, we find that entries on Vietnam exhibit all the patterns we find in our main results: all models except the Poisson model on belligerence yield a positive and statistically significant effect. However, it is also worth noting that the remaining entries on Asian countries feature the same results. Entries on Vietnam are therefore not dictating the observed patterns in racial tropes for Asian states but are reinforcing them.

The story is slightly different for the Soviet Union. Entries about the USSR are clearly associated with higher use of tropes regarding belligerence in both quasibinomial and Poisson models; these entries also feature more animal analogies in Poisson model. Notably, these effects were previously seen for the Eastern Europe variable in Table A31, but they no longer exist for Eastern Europe once the USSR is treated separately. Once the USSR is not included as an Eastern European country, we see that the region is often associated with lower use of tropes compared to other European states.

Finally, we find some evidence that South Africa is discussed using language that includes more tropes regarding irrationality, animal analogies, and belligerence. These effects do not align with the results we obtain for Sub-Saharan Africa in either Table A31 or A32. South Africa may thus be seen in a slightly different albeit still racialized manner than other countries in the region.

Table A31: Results of regional regressions on relationship between racial tropes in PDB entries and measurements of the racialized Otherness.

	Dependent variable:							
	Quasibinomial				*Poisson*			
	Infant.	Animal	Bellig.	Irrat.	Infant.	Animal	Bellig.	Irrat.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Americas	-0.773^{***}	-0.307	-0.418^{*}	0.418^{**}	0.152	-0.423^{**}	0.033	0.013
	(0.116)	(0.188)	(0.225)	(0.193)	(0.207)	(0.197)	(0.079)	(0.107)
Eastern Europe	-0.099^{***}	-0.008	0.093^{***}	0.044	-0.185^{***}	0.145^{***}	0.062^{***}	-0.044^{***}
	(0.020)	(0.026)	(0.035)	(0.033)	(0.027)	(0.030)	(0.011)	(0.012)
Middle East/Northern Africa	0.098	-0.540^{**}	-0.001	0.011	0.223^{**}	0.041	0.054	-0.300^{***}
	(0.137)	(0.235)	(0.334)	(0.203)	(0.096)	(0.125)	(0.063)	(0.042)
Sub-Saharan Africa	0.690^{***}	0.244	1.694^{***}	0.206	0.929^{***}	-4.494^{***}	0.144	0.436^{***}
	(0.200)	(0.332)	(0.470)	(0.378)	(0.205)	(1.064)	(0.104)	(0.106)
Asia	0.871^{***}	1.410^{***}	2.922^{***}	1.087^{**}	0.963^{***}	1.920^{***}	-0.265^{*}	0.298^{**}
	(0.258)	(0.462)	(0.789)	(0.480)	(0.268)	(0.595)	(0.140)	(0.122)
Oceania	-0.737^{***}	-0.132	-0.221	-0.236	-4.838^{***}	1.284^{***}	-6.045^{***}	-0.253^{***}
	(0.119)	(0.218)	(0.307)	(0.190)	(1.007)	(0.170)	(1.004)	(0.048)
Years since independence	-0.291^{***}	-0.491^{***}	-0.835^{***}	-0.362^{***}	-0.381^{***}	-0.424^{***}	0.055	-0.130^{***}
	(0.060)	(0.108)	(0.189)	(0.127)	(0.068)	(0.143)	(0.033)	(0.031)
Conflict	-0.169^{***}	0.119^{**}	-0.012	-0.219^{***}	-0.045^{*}	-0.004	0.005	0.029^{**}
	(0.053)	(0.054)	(0.034)	(0.029)	(0.026)	(0.031)	(0.019)	(0.015)
Democracy	-0.077	-0.097	-0.115	-0.152	-0.033	-0.034	0.037	-0.038
	(0.098)	(0.129)	(0.110)	(0.111)	(0.085)	(0.114)	(0.060)	(0.036)
Personalism	0.009	0.001	-0.012	-0.041	-0.040	0.104	0.049	0.0005
	(0.043)	(0.078)	(0.052)	(0.068)	(0.029)	(0.067)	(0.032)	(0.025)
Leader mention	0.128^{***}	-0.107^{***}	0.017	0.024	0.006	0.007	0.033^{***}	0.098^{***}
	(0.016)	(0.022)	(0.018)	(0.018)	(0.034)	(0.036)	(0.012)	(0.013)
US trade	-0.389	-2.641^{*}	-1.588	-0.003	-0.719	-2.274^{**}	-0.635^{**}	-0.850^{***}
	(0.689)	(1.443)	(2.123)	(1.024)	(0.523)	(1.094)	(0.283)	(0.207)
US military aid	0.016^{***}	0.007^{*}	0.028^{***}	0.008	-0.002	0.008	-0.005^{**}	0.003^{*}
	(0.004)	(0.004)	(0.005)	(0.009)	(0.005)	(0.005)	(0.002)	(0.002)
US defense	0.549^{***}	0.891^{***}	0.418^{*}	-0.243^{*}	-0.186	0.368^{***}	-0.046	0.125
	(0.067)	(0.141)	(0.216)	(0.124)	(0.138)	(0.132)	(0.042)	(0.091)
Entry length	0.289^{***}	-0.066^{***}	-0.110^{***}	0.134^{***}	0.989^{***}	0.952^{***}	1.039^{***}	1.019^{***}
	(0.013)	(0.018)	(0.020)	(0.017)	(0.024)	(0.029)	(0.008)	(0.013)
Constant	-2.371^{***}	-1.305^{***}	-0.074	-3.157^{***}	-7.095^{***}	-7.198^{***}	-5.891^{***}	-5.791^{***}
	(0.169)	(0.244)	(0.358)	(0.251)	(0.127)	(0.175)	(0.086)	(0.086)
Observations	89,016	89,016	89,016	89,016	89,016	89,016	89,016	89,016
Topics	✓	✓	✓	✓	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SEs (country)	✓	✓	✓	✓	✓	✓	✓	✓

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table A32: Results of regional regressions on relationship between racial tropes in PDB entries and measurements of the racialized Otherness, separating three key countries from their regions.

	Dependent variable:							
	Quasibinomial				*Poisson*			
	Infant.	Animal	Bellig.	Irrat.	Infant.	Animal	Bellig.	Irrat.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Americas	-0.773^{***}	-0.307	-0.418^{*}	0.418^{**}	0.152	-0.423^{**}	0.033	0.013
	(0.116)	(0.188)	(0.225)	(0.193)	(0.207)	(0.197)	(0.079)	(0.107)
Eastern Europe	-0.426^{***}	-0.855^{***}	-0.424	-0.059	-0.417^{***}	-0.285^{*}	-0.073	-0.274^{***}
	(0.105)	(0.242)	(0.334)	(0.211)	(0.110)	(0.155)	(0.057)	(0.038)
Middle East/Northern Africa	0.098	-0.540^{**}	-0.001	0.011	0.223^{**}	0.041	0.054	-0.300^{***}
	(0.137)	(0.235)	(0.334)	(0.203)	(0.096)	(0.125)	(0.063)	(0.042)
Sub-Saharan Africa	0.690^{***}	0.244	1.694^{***}	0.206	0.929^{***}	-4.494^{***}	0.144	0.436^{***}
	(0.200)	(0.332)	(0.470)	(0.378)	(0.205)	(1.064)	(0.104)	(0.106)
Asia	0.871^{***}	1.410^{***}	2.922^{***}	1.087^{**}	0.963^{***}	1.920^{***}	-0.265^{*}	0.298^{**}
	(0.258)	(0.462)	(0.789)	(0.480)	(0.268)	(0.595)	(0.140)	(0.122)
Oceania	-0.737^{***}	-0.132	-0.221	-0.236	-4.838^{***}	1.284^{***}	-6.045^{***}	-0.253^{***}
	(0.119)	(0.218)	(0.307)	(0.190)	(1.007)	(0.170)	(1.004)	(0.048)
Vietnam	0.533^{***}	1.225^{***}	2.496^{***}	1.290^{***}	0.890^{***}	1.304^{***}	-0.061	0.374^{***}
	(0.198)	(0.314)	(0.496)	(0.386)	(0.177)	(0.379)	(0.101)	(0.086)
USSR	-0.099^{***}	-0.008	0.093^{***}	0.044	-0.185^{***}	0.145^{***}	0.062^{***}	-0.044^{***}
	(0.020)	(0.026)	(0.035)	(0.033)	(0.027)	(0.030)	(0.011)	(0.012)
South Africa	0.015	-0.005	0.043	0.292^{***}	-0.470^{***}	0.418^{***}	0.197^{***}	-0.111^{*}
	(0.063)	(0.144)	(0.163)	(0.109)	(0.083)	(0.142)	(0.047)	(0.058)
Years since independence	-0.291^{***}	-0.491^{***}	-0.835^{***}	-0.362^{***}	-0.381^{***}	-0.424^{***}	0.055	-0.130^{***}
	(0.060)	(0.108)	(0.189)	(0.127)	(0.068)	(0.143)	(0.033)	(0.031)
Conflict	-0.169^{***}	0.119^{**}	-0.012	-0.219^{***}	-0.045^{*}	-0.004	0.005	0.029^{**}
	(0.053)	(0.054)	(0.034)	(0.029)	(0.026)	(0.031)	(0.019)	(0.015)
Democracy	-0.077	-0.097	-0.115	-0.152	-0.033	-0.034	0.037	-0.038
	(0.098)	(0.129)	(0.110)	(0.111)	(0.085)	(0.114)	(0.060)	(0.036)
Personalism	0.009	0.001	-0.012	-0.041	-0.040	0.104	0.049	0.0005
	(0.043)	(0.078)	(0.052)	(0.068)	(0.029)	(0.067)	(0.032)	(0.025)
Leader mention	0.128^{***}	-0.107^{***}	0.017	0.024	0.006	0.007	0.033^{***}	0.098^{***}
	(0.016)	(0.022)	(0.018)	(0.018)	(0.034)	(0.036)	(0.012)	(0.013)
US trade	-0.389	-2.641^{*}	-1.588	-0.003	-0.719	-2.274^{**}	-0.635^{**}	-0.850^{***}
	(0.689)	(1.443)	(2.123)	(1.024)	(0.523)	(1.094)	(0.283)	(0.207)
US military aid	0.016^{***}	0.007^{*}	0.028^{***}	0.008	-0.002	0.008	-0.005^{**}	0.003^{*}
	(0.004)	(0.004)	(0.005)	(0.009)	(0.005)	(0.005)	(0.002)	(0.002)
US defense	0.549^{***}	0.891^{***}	0.418^{*}	-0.243^{*}	-0.186	0.368^{***}	-0.046	0.125
	(0.067)	(0.141)	(0.216)	(0.124)	(0.138)	(0.132)	(0.042)	(0.091)
Entry length	0.289^{***}	-0.066^{***}	-0.110^{***}	0.134^{***}	0.989^{***}	0.952^{***}	1.039^{***}	1.019^{***}
	(0.013)	(0.018)	(0.020)	(0.017)	(0.024)	(0.029)	(0.008)	(0.013)
Constant	-2.371^{***}	-1.305^{***}	-0.074	-3.157^{***}	-7.095^{***}	-7.198^{***}	-5.891^{***}	-5.791^{***}
	(0.169)	(0.244)	(0.358)	(0.251)	(0.127)	(0.175)	(0.086)	(0.086)
Observations	89,016	89,016	89,016	89,016	89,016	89,016	89,016	89,016
Topics	✓	✓	✓	✓	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SEs (country)	✓	✓	✓	✓	✓	✓	✓	✓

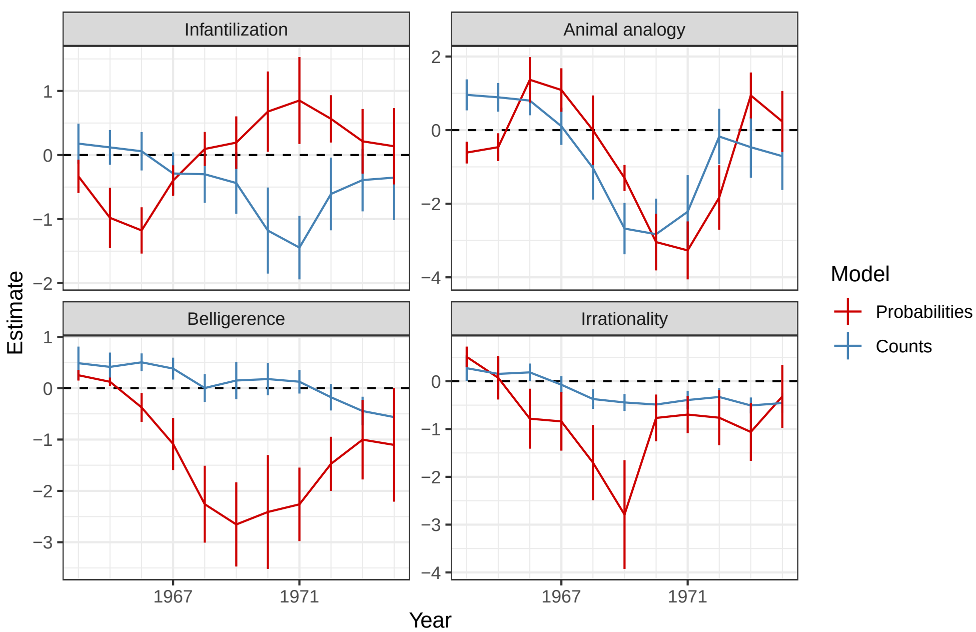
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

4.7 More on Temporal Variation

Figure A5 replicates the analysis shown in Figure 6 in the main manuscript, but replacing the *Global South* variable with the *years since independence* variable. The results for years since independence are largely a mirror image of what we find for the Global South indicator variable. Recall that we expect to see the propensity of racial tropes to decline as a country accumulates years since being decolonized; negative coefficients would be consistent with this argument. For animal analogies, belligerence, and irrationality, we see (as was the case for the Global South in Figure 6) that the most prominent effects appear for probabilistic measures of tropes derived from the supervised learning process. Moreover, many of the strongest relationships between years since independence and (decreased) use of tropes appears at the very end of the 1960s and into the 1970s.

The fact that temporal results for years since independence – a measure that is inherently linked to countries outside the United States – are highly consistent with the Global South variable provides further suggestive evidence that racial politics within in the United States were not a primary driver of changes in the prevalence of racial tropes in the PDB.

Figure A5: Coefficient estimates for the *years since independence* variable, using a moving seven-year temporal window and full models accounting for topics. Bands represent 95% confidence intervals.



References

Fatourechi, Mehrdad, Rabab K. Ward, Steven G. Mason, Jane Huggins, Alois Schlögl and Gary E. Birch. 2008. "Comparison of Evaluation Metrics in Classification Applications with Imbalanced Datasets." *Seventh International Conference on Machine Learning and Applications* pp. 777–782.

Goemans, Henk E., Kristian Skrede Gleditsch and Giacomo Chiozza. 2009. "Introducing Archigos: A Dataset of Political Leaders." *Journal of Peace Research* 46(2):269–283.

Hensel, Paul R. and Sara M. Mitchell. 2007. "The Issue Correlates of War (ICOW) Project Supplementary Data Set: Colonial History Data Set." Harvard Dataverse. `https://doi.org/10.7910/DVN/5EMETG`.

Johnson, Loch K. 2008. "Glimpses into the Gems of American Intelligence: The *President's Daily Brief* and the National Intelligence Estimate." *Intelligence and National Security* 23(3):333–370.

Kovar, Richard. 2000. "An Interview with Richard Lehman." *Studies in Intelligence* 44(3):51–63.

Le, Quoc V. and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on Machine Learning*. Vol. 32 pp. 1188–1196.

Mansfield, Celia. 2015. *The President's Daily Brief: Delivering Intelligence to Kennedy and Johnson*. Washington, DC: Central Intelligence Agency.

Mimno, David and Moontae Lee. 2014. Low-dimensional Embeddings for Interpretable Anchor-based Topic Inference. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics pp. 1319–1328.
URL: *https://aclanthology.org/D14-1138*

Priess, David. 2016. *The President's Book of Secrets: The Untold Story of Intelligence Briefings to America's Presidents*. New York: PublicAffairs.

Roberts, Margaret E., Brandon M. Stewart and Dustin Tingley. 2017. "stm: R Package for Structural Topic Models." *Journal of Statistical Software* .

Ying, Luwei, Jacob M. Montgomery and Brandon M. Stewart. 2022. "Topics, Concepts, and Measurement: A Crowdsourced Procedure for Validating Topics as Measures." *Political Analysis* 30(4):570–589.